# A Privacy-Based Mechanism for Users' Information Scoring and Anonymisation Across Multiple Online Social Networks

**Erfan Aghasian**, M.Sc.

Submitted in fulfilment of the requirements of the degree of

Doctor of Philosophy


Discipline of ICT, School of Technology, Environments and Design
THE UNIVERSITY OF TASMANIA

March 2019

# Abstract

Social network sites are becoming more and more popular among individuals in recent years and have eased social interactions to help individuals connect with others with a common interest, and to exchange information. As individuals share their personal information such as age, job details, views, opinions and thoughts on such sites, they may face different privacy issues such as identity theft, bullying, harassment and even job termination. As the participation of users in social networking sites increases, the likelihood of sharing information with unknown users escalates, and the possibility of privacy risks for the user is elevated. There are two main ways suggested in the literature to minimise the privacy risks of users on social media sites. The first is to measure the privacy risk. The second is to hide sensitive information from others. To measure the privacy risk, there are several studies on scoring privacy for online social media users for structured data (data contained in fields such as name, age and qualification) in a single source, neglecting the fact that social media users, in general, have multiple social network profiles revealing dissimilar sensitive information which can aggravate the risk. Moreover, there are limited works on privacy calculation of the risk caused by unstructured data (any textual data). For preserving the privacy of data, several anonymisation techniques have been proposed. However, in the context of preserving the privacy of individuals during the friending phase (the act of adding someone as a friend) in social media, there are only a few available approaches. Most of them disregard the privacy from the user's perspective

and are more focused on users' security rather than privacy. To address these problems, this thesis proposes approaches that can support online social network users to quantify their privacy disclosure based on their structured and unstructured information shared across multiple social media sites. Evaluation of the study illustrates that the proposed models can deliver a better approximation of privacy for users with multiple profiles on online social networks. This thesis also investigates a privacy-preserving friending method for information sharing across multiple social media sites. As friending exposes the sensitive data of a user to others, this model helps individuals to decide how to share their information safely through social networking sites with a reduced risk of being exploited or re-identified. Evaluation of the model shows that information sensitivity calculation, as well as anonymisation, offers a more effective way of friending.

The key research findings and contributions of this thesis are:

- Despite several governmental and social networks policy changes, privacy risk is still a significant problem with social media sites.
- It is important to consider the visibility and sensitivity of the information on multiple online social network sites to improve the accuracy of privacy risk evaluation.
- For accurate and credible scoring the privacy risk of unstructured data such as tweets, blogs and comments, and structured information should also be considered along with sentiment associated with unstructured data.
- By considering the sensitivity of the information in anonymisation of shared data, privacy-preserved friending can be achieved with reduced privacy risks for social media users.

# Declaration

This is to certify that this thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis is included, nor does the thesis contain any material that infringes copyright.

---

**Erfan Aghasian**, M.Sc., March 2019

# Authority of Access

This thesis is not to be made available for loan or copying for two years following the date this statement was signed. Following that time, the thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

---

**Erfan Aghasian**, M.Sc., March 2019

# Statement of Co-Authorship

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

1. Candidate = **Erfan Aghasian, School of Technology, Environments and Design, University of Tasmania**

2. Author 1, Supervisor = **Saurabh Garg, School of Technology, Environments and Design, University of Tasmania**

3. Author 2, Co-supervisor = **James Montgomery, School of Technology, Environments and Design, University of Tasmania**

4. Author 3, Co-author = **Longxiang Gao, School of Information Technology, Deakin University, Melbourne, VIC, Australia**

5. Author 4, Co-author = **Shui Yu, School of Information Technology, Deakin University, Melbourne, VIC, Australia**

**Author details and their roles:**

**Paper 1.** "User's Privacy in Recommendation Systems Applying Online Social Network Data, A Survey and Taxonomy", ***Big Data Recommender Systems: Recent Trends and Advances, The Institution of Engineering and Technology (IET), Book Chapter.***

Located in Chapter 2

Candidate was the primary author of the paper and contributed 80% in preparing the paper. The contribution of authors 1 and 2 was individually 10%.

**Paper 2.** "Scoring users' privacy disclosure across multiple online social networks", ***IEEE ACCESS.***

Located in Chapter 3

Candidate was the primary author of the paper and contributed 60% in writing, simulation, analysing the proposed method and preparing the paper. The contribution of authors 1 and 2 was individually 15% and the contribution of authors 3 and 4 was individually 5%.

**Paper 3.** "A Privacy-Enhanced Friending Approach for Users on Multiple Online Social Networks", ***Computers.***

Located in Chapter 5

Candidate was the primary author of the paper and contributed 70% in writing, simulation, analysing the proposed method and preparing the paper. The contribution of authors 1 and 2 was individually 15%.

**Paper 4.** "An Automated Model to Score the Privacy of Unstructured Information - Social Media Case", ***Computers & Security, Submitted***

Located in Chapter 4

Candidate was the primary author of the paper and contributed 70% in writing, simulation, analysing the proposed method and preparing the paper. The contribution of authors 1 and 2 was individually 15%.

Signed: ——— —

| | |
|---|---|
| Dr. Saurabh Garg | Prof. Mark Hunt |
| Supervisor, ICT Discipline | Head of School |
| University of Tasmania | University of Tasmania |
| Date: 23/01/2019 | 25/01/2019 |

# Ethics of Ethical Conduct

The research associated with this thesis abides by the rules of the Tasmania social science human research ethics committee (HREC) containing minimal risk ethics application approvals with the ethics references codes H0016696 and H0016693.

# Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors Dr Saurabh Garg and Dr James Montgomery for their continuous support of my Ph.D. study and related research, for their patience, motivation, and immense knowledge. Their guidance helped me at all times during this research. I could not have imagined having better supervisors and mentors for my Ph.D. study. Besides my supervisors, I would like to thank Dr Longxiang Gao and Dr Shui Yu from Deakin University for their insightful comments during my Ph.D. research. I thank all the staff of the ICT department of the University of Tasmania in helping me and for their support and for their friendships during my Ph.D. program. I would also thank Jill Smith who proofread the thesis. Last but not the least, I would like to thank my beloved parents, my siblings and my friends for supporting me spiritually during the writing of this thesis and in my life in general. They are the most important people in my world and I dedicate this thesis to them.

# Contents

# List of Figures

# List of Tables

# Abbreviations

The following list describes the various abbreviations used throughout the thesis.

- OSN = Online Social Networks
- UGC = User-generated Content
- Infosec = Information Security
- PISX = Privacy Index
- IRT = Item Response Theory
- PII = Personally Identifiable Information
- QI = Quasi-Identifier
- PDS = Privacy Disclosure Score
- FIS = Fuzzy Inference System
- LV = Linguistic Variable
- MF = Membership Function
- FB = Facebook
- RG = ResearchGate
- LD = LinkedIn

- G+ = Google Plus
- LIWC = Linguistic Inquiry Word Count
- ML = Machine Learning
- NLP = Natural Langage Processing
- CRF = Conditional Random Fields
- SVM = Support Vector Machine
- POS = Part Of Speech
- ABE = Attribute-based Encryption

# Chapter 1
# Introduction

*"With Social Media so prevalent we are all extremely visible. Your prospective clients, your peers and your competition can drill as deep as they wish searching, reading and gathering information online about you and posted by you without you ever knowing who's searching. Depending on what they find, your prospects may choose to do business with you or not." (Mari Smith)*

Online social networks (OSNs) have become an integral part of individuals' everyday life and have changed the way individuals connect (Guo et al., 2015). As an example, the growth rate of Facebook has been reported as being as high as 3% per week (Wondracek et al., 2010), while 1.47 billion users were active daily in the second quarter of 2018 (Fulgoni, 2018). There are different social network sites which provide different features for their users. Social media technologies have been classified based on how individuals interact with each other. For instance, social media sites such as Facebook, are primarily used to share updates of the daily encounters of individuals, as they occur, particularly photos. LinkedIn is mainly used to share career information and qualification information for job seekers and recruiters. Wikis are normally websites formed to deliver educational information (Osatuyi, 2013).

In such sites, individuals have the ability to create content, share it with others and discuss it with several individuals; the volume of this information is increasing (Cheung et al., 2011). Users are allowed to share both structured and unstructured information such as age, thoughts, posts, photos and videos, updates

on their daily activities and events and other user-generated content (UGC) with other individuals in their network (Obar and Wildman, 2015).[1]

Obviously, not all information is meant for disclosure to everyone. Although a considerable part of this data is not sensitive, it is not unusual for individuals to share some sensitive data as well, such as current town or marital status, on social networking sites (Walden, 2002). As users create different social profiles on different platforms, they provide more sensitive information which would be accessible to unknown users as well and this may create privacy issues for them. Figure 1.1 shows how different pieces of information can be aggregated from a user's social profiles. For example, a user may disclose his background information on Facebook while he/she may reveal his/her job information in the LinkedIn account.

Privacy issues include spamming, unauthorised access to information, online bullying, harassment, trolling, relationship breakups, job termination, social anxiety, social overload and personal relationship problems, all of which can have negative impacts on individuals' lives (Zheleva et al., 2012).

Hence, there is a need to increase users' awareness of privacy-related issues and preserve their privacy while they share their information in such sites. Normally, online social networks are capable of protecting users' privacy via confidentiality agreements, but they are not sufficient to protect users in the face of the privacy leakages that occur every day (Kafalı et al., 2014). As the social network sites rely on the quantity of personal and probably sensitive information that they make accessible, there is a continuous threat that an adversary can aggregate data or manipulate the structure of a social network to deduce information about recorded users, even when these sites engage mechanisms to guard the secrecy of their users (BACKSTROM et al., 2007; Chew et al., 2008).

[1] In this context, structured data refers to information that can be displayed in columns and rows and can easily be ordered and processed by data mining tools. In contrast, unstructured data refers to raw and unorganised information such as the content of a tweet or a post in social media.

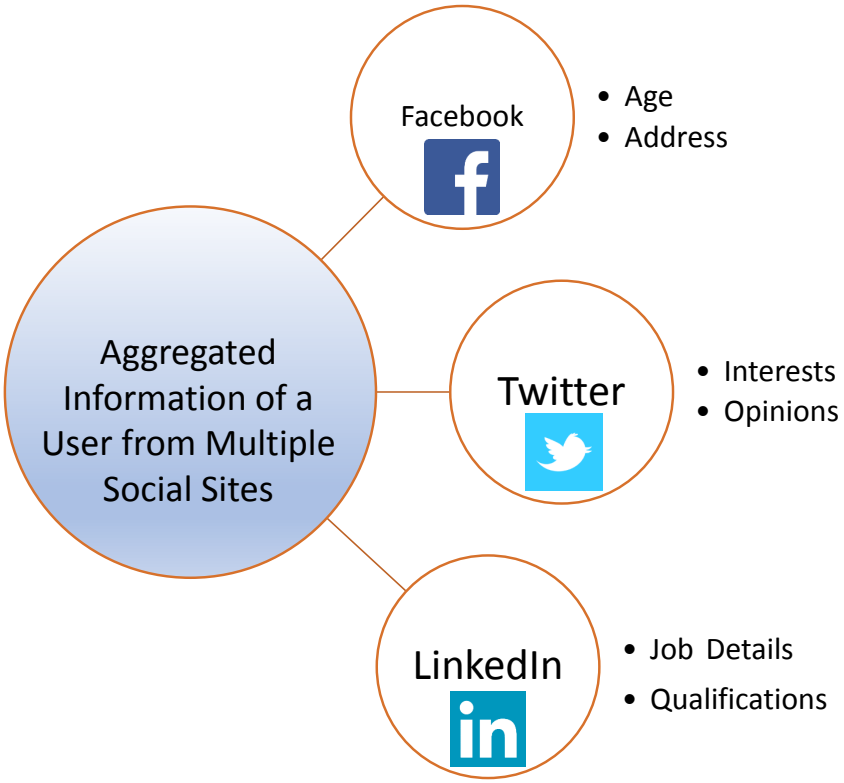



Figure 1.1: Information Aggravation from Multiple Social Sites

To protect the users' online privacy before they encounter any serious problems in their daily lives, there are mainly two suggested ways. The first is to assess the privacy of users' information (Liu and Terzi, 2010a). Measuring their privacy can help users to understand the causes of privacy risks from the shared information (Tuunainen et al., 2009). The second is to hide individuals' sensitive information from other users and adversaries (Zheleva and Getoor, 2009). While scoring privacy provides an insight for users regarding their privacy levels, information hiding can help users to share their information in such a way that the least possible amount of risk exists for them in case of information disclosure.

Although there are various studies for scoring and preserving the privacy of users in online social media, they only focus on a single source of data. Hence, these methods may not be satisfactory for users who usually have multiple online social profiles, disclosing different sensitive information which can aggravate the privacy risk. Accordingly, there is a need for models which can quantify the privacy risk from multiple social media platforms. This research aims to provide frameworks for scoring privacy and for preserving the privacy of the users in terms of friending to reduce the vulnerability of their information, while also trying to increase the awareness of the users who are involved in online social network sites.

This study would be beneficial to all social network users and individuals for preventing their sensitive information being compromised. It is an expectation that this would increase the awareness of users of security issues and possible threats to their privacy. This is a comparatively novel viewpoint in the study of privacy measurement and privacy-preserving friending, and one that is likely to create valuable supplementary insights for the future. From a practice standpoint, this study is relevant to the social networks user. For future researchers, the study can provide reference information on the recent status of users' privacy issues in public, as well as available data sources and countermeasures for privacy

preserving.

The rest of this chapter is structured as follows: firstly, a short introduction is given to the concept of users' privacy risks and concerns in social networks and data publishing. Secondly, the objectives of the research are presented in order to highlight the importance and significance of undertaking the project. Then, the proposed methodology and solution for the defined objectives will be outlined. Next, the contributions of the thesis will be explained. Finally, an outline of the research design and thesis structure will be presented.

## 1.1 Problem Statement and Research Questions

This thesis focuses on the following problem:

*How to minimise the privacy risks for online social media users who reveal their information in multiple online social networking sites.*

In comparison with traditional systems, the privacy objectives of online social networks are slightly different. In this perspective, users' data and identity should be protected against unauthorized access and modification to retain integrity. Users should be aware of the privacy risks which they could encounter while sharing their information. As users share their information on multiple online social networking sites, it becomes significantly more difficult for users to control and classify their shared information with their friends and other parties, in order to avoid any privacy breaches. Regarding this fact, the research questions of this study are:

1. How are privacy issues currently addressed in different social networking sites and what privacy threats may users face when they share their information on multiple online social networking sites?

2. How should the privacy score of shared information (both structured and

unstructured data) of each user be measured in multiple online social sites?

3. How can the privacy of users be preserved while they share their information with other individuals in terms of friending?

## 1.2 Research Objectives

The first aim of this study is to conduct a comprehensive investigation of privacy risks and threats that users may face on online social network sites and also to investigate the current scoring and preservation methods. The second aim is to present a new framework for measuring the privacy score of users who share information on multiple online social networks. The proposed model is applied to structured information which is shared by social media users. By applying this framework, users can understand their privacy level compared with other users. Thereafter, a novel model for scoring the privacy of unstructured information (free-format text) is presented which can warn users about the level of risks of the shared contextual information in their social network profiles. At the end, a novel mechanism for friending, with a reduced risk is proposed for preserving and enhancing users' privacy. Hence, the focus of this thesis is on proposing ways to mitigate the privacy risks of users who share their information in such sites. Here, the focus is not on what benefits social media sites can bring for users or on what their functionalities are. The objectives of this research are to:

1. Perform a comprehensive study of theoretical and empirical insights into user's privacy and privacy-preserving perspective in privacy measurement and friending.

2. Propose models for measuring the users' privacy score for structured and unstructured information and to build the awareness of users regarding sharing of sensitive information.

3. Propose a novel technique to provide friending with reduced risk among online social network users considering privacy-preserving techniques for structured data.

It is hypothesised that *with the aggregation of a user's information from multiple sources of social networking sites, a more precise measurement of individual's privacy risks can be achieved*.

## 1.3 Proposed Solution

As the nature of the data and the social media sites varies, the control and classification of sensitive information is becoming more complicated for users. Hence, there is a need for improvement of the privacy of users who participate in online social media sites, as well as a need for the proposal of mechanisms to keep users aware of their privacy risk level while they participate in multiple networks. Moreover, there is a need for mechanisms to keep users' sensitive information preserved from various attacks, such as information exploitation, impersonation and so on.

There are several studies (Liu and Terzi, 2010a; Srivastava and Geethakumari, 2013; Veiga and Eickhoff, 2016; Wilson et al., 2005) about scoring the privacy of shared structured data (any data that resides in a fixed field within a record or file such as relationship status, age and home town) by focusing on a single source of data. In order to measure the privacy risk of users in social media for the structured data, two factors should come into consideration, namely, visibility and sensitivity of the information. Visibility describes how a piece of information can be seen by others and sensitivity defines how valuable that piece of information is valuable for that user. By combining these factors, a more accurate estimation of the privacy score can be obtained which helps users understand their privacy level, as compared with other users.

Further, there are limited studies for measuring the privacy risks of shared unstructured data (data that cannot be so readily classified and added to a field, such as photos). Although authors (Farzindar and Inkpen, 2015; Wilson et al., 2005) have proposed methods to measure the amount of information leakage from a social networking sites, most of the studies only focus on the polarity of the sentiments regardless of the privacy risk level for each sentiment. To achieve this, there is a need for a framework which can calculate the privacy risks of shared unstructured data. To measure the privacy of the unstructured data, it is critical to comprehend the factors which impact on the sentiments' privacy. To obtain the privacy-related factors, different machine learning algorithms have been applied and, then, have been added to a fuzzy system in order to measure the final privacy risk accompanied by the number of negative and positive terms of each sentiment. The evaluation of the model indicates the effectiveness of privacy scoring for unstructured data while users share different information on multiple online social networking sites.

For preserving the privacy of users, several anonymisation techniques have been proposed (Guha et al., 2008; Baden et al., 2009; Domingo-Ferrer et al., 2008). While most of the works focus on users' security, rather than the privacy, in the terms of friending in online social networks, there is a need to develop approaches which can satisfy the online privacy of individuals in such networks. Hence, this research strives to close the gap of friending in multiple sources of data and proposes a privacy-based model to solve the problem of the identified gap. To achieve this, the sensitivity calculation of the shared information has been considered as the main factor which impacts on information sharing. Then, related formulae have been proposed to calculate the sensitivity of the information users share in such sites. Assessment of the model indicates that hiding sensitive information on social media can reduce the risk of re-identification or exploitation in such sites and can guarantee the users privacy in a more efficient way.

## 1.4 Thesis Contributions

This thesis contributes towards the improvement of the privacy of users who participate in multiple online social sites considering both structured and unstructured information. The contributions are as follows:

1. This thesis provides a comprehensive taxonomy of user's privacy in online social media and recommender systems that covers various aspects such as privacy concerns in social networks, user-specific risks, privacy measurement techniques, privacy-preserving approaches and privacy preservation models. The taxonomy is intended to provide researchers with a detailed view of the goal of privacy measurement and preservation, by providing insights to key issues that are still outstanding. The taxonomy and survey also highlight various research gaps to enhance the online social media users' privacy.

2. This thesis presents the design and development of privacy scoring frameworks for both shared structured and unstructured information which provide privacy awareness for online social media users on multiple social sites. It highlights the benefits of privacy scoring from multiple social sites compared with a single source of information and provides insights for users about the level of their online privacy.

3. This thesis models a privacy-preserved friending mechanism for online social media users. By considering the advantages of information sensitivity calculation and anonymisation algorithms, a privacy preserved model with reduced risk which can help users to share their information in a safe manner is proposed.

## 1.5 Thesis Structure

The rest of the thesis is organised as follows: Chapter 2 presents the survey and taxonomy of users' privacy on online social networks and recommender systems. This chapter offers the literature background for the remaining parts of this thesis by highlighting research gaps in privacy scoring and privacy-preserved friending. Chapter 3 describes in detail a privacy scoring model for shared structured data, for online social media users. In this chapter, the proposed model is compared with the most recent proposed models and is evaluated to show its benefits. Chapter 4 proposes a privacy scoring model for social media users in the context of the unstructured data. Chapter 5 presents the privacy-enhanced friending model to enable users to connect with others, but with reduced risk. Chapter 6 concludes and offers directions for future work.

Chapter 2 has been removed for copyright or proprietary reasons.

It has been published as: Aghasian, E., Garg, S., Montgomery, J. User's privacy in recommendation systems applying online social network data, a survey and taxonomy, in, Big data recommender systems - Volume 1: Algorithms, architectures, big data, security and trust , The Institution of Engineering and Technology (IET), 2019, Khalid, O., Khan, S. E., Zomaya, A. Y. (eds.)

# Chapter 2

# Users' Privacy in Online Social Network Data

# Chapter 3
# Scoring Users' Privacy Disclosure on Multiple Online Social Networks

*"My take is, privacy is precious. I think privacy is the last true luxury. To be able to live your life as you choose without having everyone comment on it or know about it." (Valerie Plame)*

In this chapter, an approach that can help social media users to measure their privacy disclosure score (PDS) based on the information shared across multiple social networking sites is investigated. Here, structured data for privacy measurement is considered. In particular, the main factors that have an impact on users' privacy, namely, sensitivity and visibility are identified to obtain the final disclosure score for each user. By applying the statistical and fuzzy systems, the potential information loss for a user can be specified by using the obtained PDS. The evaluation results with real social media data show that this method can provide a better estimation of privacy disclosure score for users having presence in multiple online social networks.

## 3.1 Introduction

Online social media users should have adequate awareness of their privacy and know the risks they may encounter by sharing their information online. Users should also be able to protect their sensitive information from their relatives,

neighbours and anyone else with whom they have shared their information with and still maintain their secrecy (Wittes and Kohse, 2017). In general, it is not easy to estimate the privacy risks from the shared information in social media. Although there are several privacy settings in social media that can be applied by users, these settings are often complex and time-consuming to adjust; most users feel confused about them and typically ignore or skip them (Zheleva et al., 2012). Hence, there is a need to have a model for automatic quantification of privacy risks to create a better view of information revelation for users. By applying a scoring framework and privacy awareness enhancing models, individuals can have a better scheme of their privacy and apply security procedures to increase their level of privacy if needed.

Several attempts (Srivastava and Geethakumari, 2013; Liu and Terzi, 2010a) have been made to quantify the privacy of a user, although most of them are designed to consider only one source of information. These may not be sufficient as each user may have multiple social network accounts for different purposes. One source of data may not disclose a wide range of information of a user that can pose a privacy risk, but when this information is combined with information from different sources, it can be risky and dangerous. Veiga and Eickhoff (2016) showed that there is an increase in privacy leakage due a user having multiple online social network platforms compared with a single source. For example, a user normally shares his/her personal information in Facebook which may pose a privacy risk. This user may share his/her occupation history and background on another site such as LinkedIn. His/her job information has again its own privacy risk, but a combination of the information from the two social media accounts can expose the user to higher risk as more information is revealed. Consequently, by considering the overall information from multiple sources, a more accurate quantification of the privacy disclosure score can be obtained.

To quantify the privacy risk of a user, a scoring function is proposed. The

inputs to this scoring function consider a set of common personal attributes that may be discovered through social networking sites. The explicit privacy settings for each item and their frequency of occurrence, both within and across social networking sites, are all considered as inputs to the privacy scoring computation in this model. In this work the factors that have an impact on the privacy of the user (sensitivity and visibility of information) are analysed.

For each factor, a comprehensive explanation of how to calculate that factor is provided and then the way to measure the final privacy disclosure score that is related to these two factors is described. If more than one source of online social network data set is being considered, each attribute of a user has different states of visibility. Hence, due to the complexity of dependency between these inputs, formulating a single formula is not trivial. Thus, fuzzy-based methods were proposed to design the model. The solution for a specific social network site was not limited but the proposed model can be deployed in all social networks. Moreover, users can be informed about their privacy level and how much data they have been shared in such networks.

The next section formulates the problem. Section 3.3 specifies the design and the mathematical formulation for calculating the privacy disclosure score. Section 3.4 presents the evaluation of the model using real social network data. Section 3.5 discusses the comparison of the study with other techniques. The final section presents the summary of this work.

## 3.2 Problem Definition

In measuring users' privacy in online social networks, two general factors can be treated as inputs for measuring the privacy disclosure score of users; these are visibility and sensitivity of information. While calculating each factor is a difficult task, this issue becomes more significant where there are multiple sources of

data and users reveal their attributes and information on different sites. These attributes and information can be either structured or unstructured data.

In the first step of this research, it is focused on answering the following questions:

- What factors influence on users' privacy in online social networks?
- How can the privacy disclosure of each user for the shared structured information in multiple social media profiles be measured?

Since the privacy disclosure score of a user can be measured, users can understand what is their privacy level compared with other users. Thus, users can pay more attention to their privacy to bring it to an acceptable level. For measuring the privacy disclosure score, it is considered that users' attributes (such as contact numbers, emails, addresses, job details, hobbies and interests) can be gained from $n$ different sources. For calculating the privacy disclosure score, the sensitivity and visibility of information as the inputs for the proposed system was measured. Function ($F_{sen}$) indicates the sensitivity of each attribute of the users from multiple sources.

Beside calculating the sensitivity, formulas to calculate the factors that have impact on users' visibility need to be provided. These factors are known as accessibility to information ($F_{acc}$), difficulty of data extraction of users' information ($F_{dif}$) and the data reliability for each attribute ($F_{rel}$).

For a user, the score of privacy is calculated as a combination of the sensitivity score and the visibility score of the user combining several attributes such as name, age, gender, email, hometown, job details and interests from different data sources. Here, it was assumed that each user is involved in multiple social networks and each attribute is disclosed to the other users in different manners, based on the usage of the social network. For example, the job details on a social network site like LinkedIn are probably more visible than on another social

network site such as Facebook.

## 3.3 Privacy Scoring Framework

Figure 3.1 presents the overview of the privacy disclosure score framework. In the first phase, the attributes for calculating the privacy are considered. These attributes can be extracted from structured data (such as username, family or Age) or obtained from unstructured data (such as blogs, messages and images). It should be noted that this research is not concerned with the technologies that can extract these attributes or the methods that can be used to collect the data. After obtaining the framework attributes, the sensitivity and the visibility of users is computed. At first, the sensitivity of the information is measured. It is taken into account that some attributes like religious and political views are more sensitive than others. These factors are to be considered in computation of the sensitivity.

Next, calculation is made on the visibility, based on three factors that have a direct impact on visibility (accessibility to information, the difficulty of data extraction and reliability of data). The overall privacy disclosure score is finally obtained from the combination of sensitivity and visibility scores. Finally, the result is analysed and the users are informed as to how strong their privacy level is, in comparison with other users.

### 3.3.1 Calculation of Sensitivity

Sensitivity shows the risk associated with the attributes of the user. When the sensitivity of an attribute increases, the risk posed by information disclosure of the individuals also increases. Srivastava and Geethakumari (2013) calculated the sensitivity score for 11 attributes for measuring the privacy score based on the quotient model. Their results indicated that the most sensitive attributes are

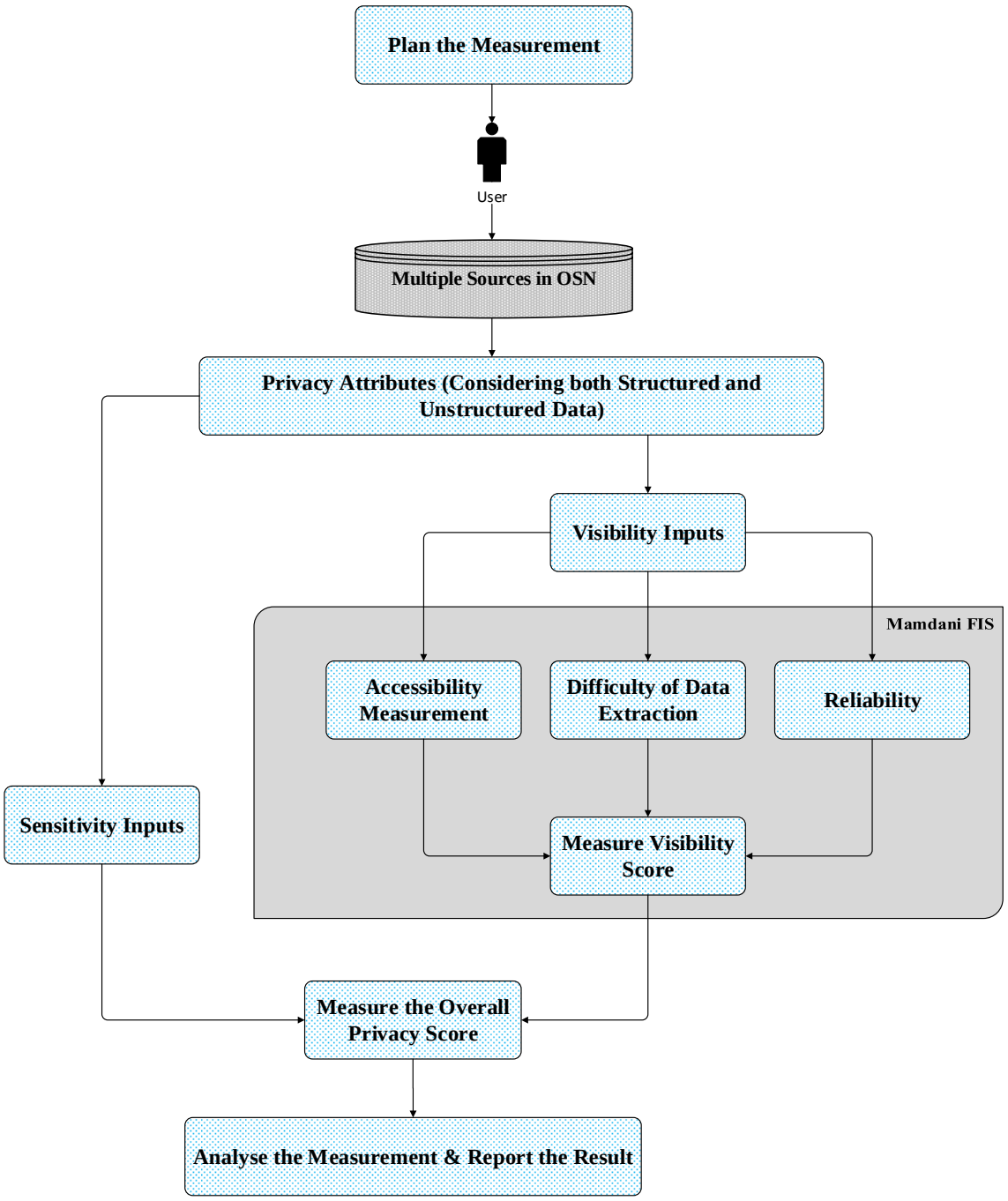



Figure 3.1: Overview of Privacy Score Framework

related to political views, religious views, contact number and relationship status. In contrast, birthdate and current town details are not that sensitive for the users. For sensitivity score ($F_{Sen}$), the sensitive values derived by Srivastava and Geethakumari (2013) are employed. This is shown in Table 3.1 for each profile item in their sample. The range of the sensitivity is between 0 and 1 where 0 indicates the lowest sensitivity and 1 indicates the highest.

Table 3.1: Sensitivity score for users' attributes

| Attributes | Sensitivity |
|---|---|
| Contact Number | 0.6 |
| E-Mail | 0.1833 |
| Address | 0.85 |
| Birthdate | 0.1166 |
| Hometown | 0.15 |
| Current town | 0.1166 |
| Job Details | 0.2 |
| Relationship Status | 0.4166 |
| Interests | 0.3 |
| Religious Views | 0.5666 |
| Political Views | 0.6833 |

### 3.3.2 Calculation of Visibility

Visibility determines how widely accessible the attributes of a user are in an online social network. For calculating the visibility, three factors that influence the visibility of user information on online social networks were considered. These factors are ease of accessibility, the difficulty of data extraction and the frequency of occurrence of information disclosure. The current predefined permissions for attributes satisfy the visibility of each item for each user. While some information of a number of users is publicly available, other attributes can be private or semi-private.

### *Accessibility Calculation*

Accessibility is defined as a measure of permissions that are given for sharing information with others. In other words, accessibility indicates how many people can have access to a specific piece of information and to what level. There are four different levels for users' information accessibility. The information can be (a) accessible only by the owner of the information, (b) accessible by his/her friends, (c) accessible by his/her friends of friends and finally (d) publicly available.

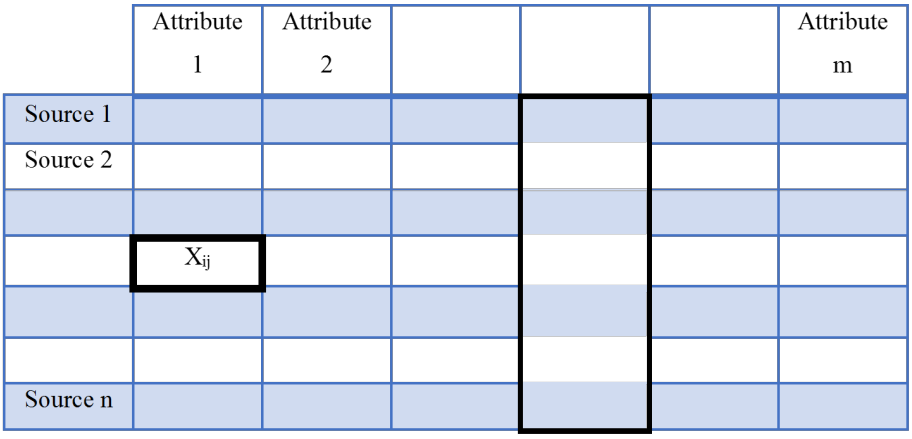



Figure 3.2: Structure of accessibility score matrix for privacy measurement for each user

An Accessibility Value $(AV)$ between 1 and 4 is given to each attribute ($1 \rightarrow$ not accessible except data owner, $2\rightarrow$ accessible by friends, $3\rightarrow$ accessible by friends of friends, $4\rightarrow$ publicly available). For calculating the accessibility of each profile item ($F_{acc}$), it is assumed that each user is participating in $n$ different online social network sites while they have different sensitive attributes. The sources indicate in which online social network a user participates. Based on the nature of each online social network, accessibility values may vary. As a case in point, the interest of a user can be accessible much more easily than his/her educational network, such as academia. Let $i$ be source, $n$ be the number of sources, $j$ be an attribute, and $m$ be the total number of attributes. Figure 3.2 shows the structure of

accessibility score matrix for calculations. After assigning the accessibility values to each of the attributes for each social network, an algorithm ($F_{acc}$) to calculate the accessibility is provided (Refer to algorithm 1). After initialising the matrix, the *temp* function sums up the total accessibility value of each profile item of a user from the source he/she has shared the information.

**Algorithm 1:** computing accessibility for an attribute

```
Data: Input: User response matrix with m columns and n rows
Result: Accessibility score of each attribute;
// Initialize the temp matrix;
for k = 1 : m do
    // Extract the k^th column and put it in col variable
    Based on the input, delete the entries that do not meet the condition i.e.
     if the difference between the maximum number and minimum number
     in each column is equal to three, delete ones;
    // calculate the mean after checking the defined
       condition
end
for j = 1 : m do
    Initiate counter and sum variables with value equal to zero;
    for i = 1 : n do
        if temp(i, j) ≠ 0 then
            sum = sum + input(i,j);
            counter = counter + 1;
        end
        Display the calculated accessibility values;
    end
    means(1,j) = sum/counter;
end
```

It should be noted that the reason for deleting '$x_{ij}$' when the range is equal to 3 is that there is an attribute publicly available in one source, while its accessibility is completely private in another source(s). While user data in a source is publicly available and anyone can have access to it, user's "only accessibility" preference does not make sense in other sources. Therefore, it can be concluded that the data that have an impact on user privacy should be calculated. In another scenario, the

data accessibility might not be publicly accessible or could only be accessed by the user. The privacy measurement can be calculated by the by users' defined permission to the information. In this case, the mean of the accessibility value of each attribute for each user is calculated.

### *Data Extraction Difficulty*

One factor that is important to compute the privacy is the difficulty of extracting private information from different formats of data. Extracting attributes from structured data is much easier than from unstructured data. For example, it is harder to understand a user's religious view from his/her picture, than from the network in which he/she clearly stated religious views. For calculation of the difficulty, three levels have been defined ($3 \rightarrow$ low difficulty, $2 \rightarrow$ medium difficulty, $1\rightarrow$ high difficulty). Naturally, the more a profile item is accessible; the less difficult data gathering is. To compute how difficult it is to extract an attribute, the mean of extraction difficulty of each attribute for each of the social networks is calculated.

$$F_{dif} = \sum_{n}(dif_j)$$

where $dif_j$ indicates the difficulty of extracting an attribute from a social network data.

### *Data Reliability Calculation*

Reliability is a criterion that can describe with what confidence a particular attribute has been disclosed in one or multiple sources. In this context, for each attribute of a user, the overall reliability of data disclosure for each attribute of the users is considered in order to appraise it in a total visibility calculation. As reliability of sensitive information will increase with a greater number of resources

validating it, a sigmoid function to measure the reliability of data is used. The reason for using the sigmoid function is that this function supports the differentiation of the reliability. The equation for calculating the reliability given by:

$$F_{rel} = \frac{2}{1 + e^{-s}} - 1$$

where 's' indicates the number of sources in which the attribute has been revealed. The output boundary for this function is [0,1], where the number of sources of disclosure increases, the reliability increase.

### *Total Visibility Calculation*

The proposed method for calculating the overall visibility score for the users is based on a set of fuzzy rules that occurs for users in different situations. The reason for choosing the fuzzy inference system (FIS) (Ghanei and Faez, 2016; Grabisch et al., 2013) is based on the nature of the system and the process complexity, and this involves various interacting parameters. Hence, FIS is considered to be a suitable method for application to this type of decision system. After defining the rules based on the inputs (Calculated numbers for the accessibility, difficulty of data extraction and frequency of occurrence), the Mamdani fuzzy inference (Wang et al., 2013) is used. Assume that a user wants to know what is his/her visibility level if his/her personal information is revealed in multiple datasets. The designed fuzzy system can explain to the user at which level of privacy (in context of visibility) he/she is situated. The process of FIS (Figure 3.3) based on Mamdani's method (Wang et al., 2013), would be as:

1. **Fuzzification (of inputs):** antecedent evaluation for each rule – obtain membership values from crisp values

2. **Implication:** obtain the consequences of each rule

3. **Aggregation:** combining step 2 output for each rule into a single fuzzy set by using a fuzzy aggregation operator

4. **Defuzzification:** obtain a crisp number as the output

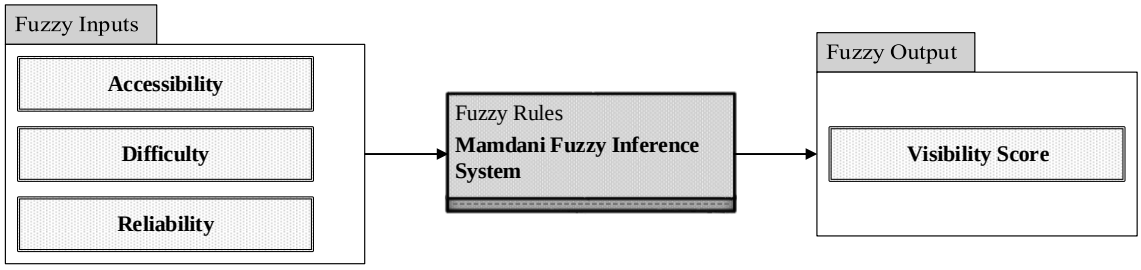



Figure 3.3: Fuzzy Inference System Overview

In the fuzzification step, a generalised bell function was selected as the membership function to define the fuzzy sets. The generalised bell function is given by:

$$f(x,[a,b,c]) = \frac{1}{1+|\frac{(x-c)}{/}a|^{2b})}$$

As can be seen, this function depends on three parameters $a$, $b$ and $c$. Each of these parameters has a physical meaning. Parameter $c$ determines the centre of the corresponding membership function. Parameter $a$ is the half width; and $b$ controls the slope at the crossover points. Figure 3.4 shows an example of a plotted generalised bell function. Table 3.2 presents the details of each membership function used in fuzzy inference model.

Apart from the membership function details, a set of rules was defined to make a logical calculation for the visibility of a user's attributes based on the inputs (Table 3.3 - the notations in the table are as follows: VL= Very low, L= Low, M=Medium, H=High, VH=Very High, X=Can be in any state). According to the fuzzy inference system model and fuzzy logic, the logical $AND$ operator were treated as 'min' while the $OR$ operator treated as a 'max' operation on the cor-

Table 3.2: Membership Function Database

| LV | Type | MF | Range | a | b | c |
|---|---|---|---|---|---|---|
| Accessibility | Input | Very Low | [0,4] | 1.171 | 11.8 | 0 |
| Accessibility | Input | Low | [0,4] | 0.449 | 7.73 | 1.85 |
| Accessibility | Input | Medium | [0,4] | 0.367 | 5.439 | 2.67 |
| Accessibility | Input | High | [0,4] | 0.667 | 14.28 | 4 |
| Difficulty | Input | Low | [0,3] | 1.16 | 14.48 | 0 |
| Difficulty | Input | Medium | [0,3] | 0.492 | 6.67 | 1.81 |
| Difficulty | Input | High | [0,3] | 0.5 | 8.56 | 3 |
| Reliability | Input | Very Low | [0,1] | 0.172 | 18.1 | 0.25 |
| Reliability | Input | Low | [0,1] | 0.136 | 19.2 | 0.567 |
| Reliability | Input | Medium | [0,1] | 0.086 | 13.2 | 0.798 |
| Reliability | Input | High | [0,1] | 0.02 | 12.3 | 0.919 |
| Reliability | Input | Very High | [0,1] | 0.132 | 32 | 1.08 |
| Visibility | Output | Very Low | [0,8] | 1 | 11.2 | 0.799 |
| Visibility | Output | Low | [0,8] | 0.628 | 7.59 | 2.64 |
| Visibility | Output | Medium | [0,8] | 0.792 | 23.6 | 11.4 |
| Visibility | Output | High | [0,8] | 0.688 | 12.3 | 5.69 |
| Visibility | Output | Very High | [0,8] | 1 | 11.1 | 7.608 |

Linguistic Variable (LV), Membership Function (MF)

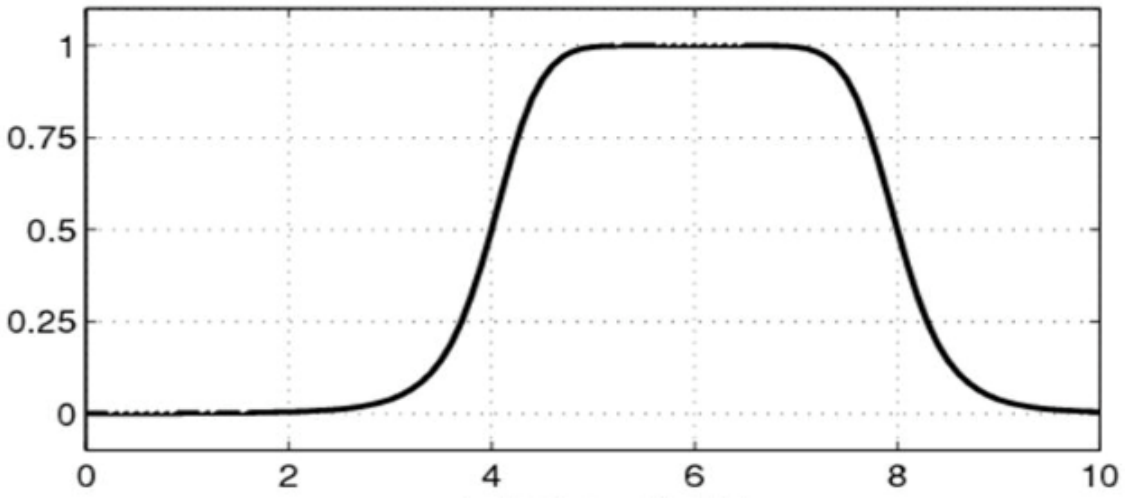



Figure 3.4: An Example of Generalized Bell Function

responding membership function. Thereafter, *max* operation for the aggregation of the database rules on the resulting of the resultant (corresponding) rules was applied.

The fuzzy rules in the model have been obtained after several consultations with experts in the domain knowledge. The membership functions are defined such that they most precisely match the values of a particular attribute. For example, if the accessibility of data is high (that is, the data is publicly available), and the frequency of occurrence is high (data published in more than three sources) as well, then the visibility (regardless of the value of data extraction difficulty) is high. The rest of the rules have been defined by this method (applying the experts knowledge).

The last step in a fuzzy inference system is defuzzification. A defuzzification method permits the acquisition of a crisp number from a fuzzy value. The two most practical methods are: mean of maxima and centroid (centre of mass) (Torres-García et al., 2016). In this model, the centroid function is exploited (which provide us wish a better result compared with other fuzzy functions), which indicates the centre of the area under the curve to obtain a crisp value for the output (visibility). This method computes the output (a crisp number) from defined rules

Table 3.3: Fuzzy Rules Database

| | Inputs | | | Output |
|---|---|---|---|---|
| | Accessibility | Difficulty of data extraction | Frequency of occurrence | Visibility |
| 1 | X | X | VH | VH |
| 2 | L | L | VL | VL |
| 3 | L | M | VL | VL |
| 4 | L | H | VL | L |
| 5 | VL | X | VL | VL |
| 6 | VL | L | M | L |
| 7 | VL | M | M | M |
| 8 | VL | H | M | M |
| 9 | VL | X | L | VL |
| 10 | VL | X | H | M |
| 11 | L | X | L | L |
| 12 | L | X | M | M |
| 13 | L | X | H | M |
| 14 | M | X | VL | L |
| 15 | M | X | L | M |
| 16 | M | L | M | M |
| 17 | M | L | H | H |
| 18 | M | M | H | H |
| 19 | M | H | H | VH |
| 20 | H | X | VL | L |
| 21 | H | L | L | L |
| 22 | H | M | L | L |
| 23 | H | H | L | M |
| 24 | H | X | M | M |
| 25 | H | X | H | H |

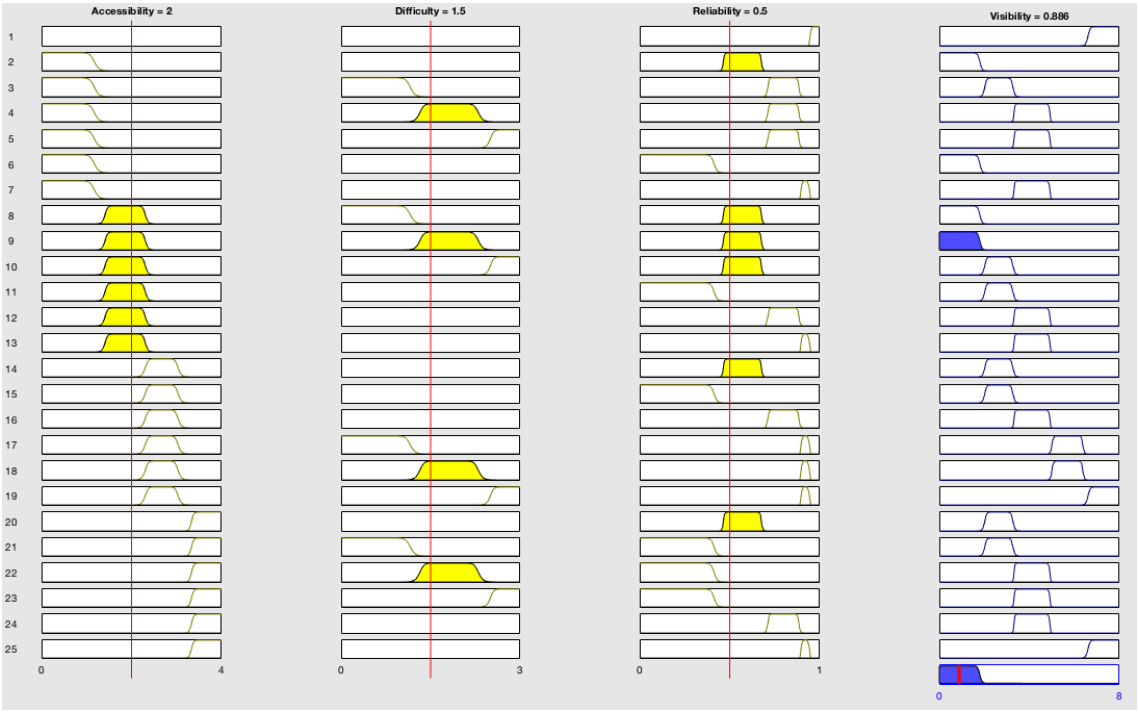



Figure 3.5: FIS model for visibility score calculation

(as input) as:

$$F_{vis}(x) = \frac{\int_{x1}^{x2} xf(x)dx}{\int_{x1}^{x2} f(x)dx}$$

where the centroid function of the area bounded by $B = [x1,\ x2]$ and the x-axis, and the function $F_{vis}(x)$ converts points of 'B' to a crisp value. The obtained value can be considered as the visibility score for the users' attributes. The FIS model for a sample is illustrated in Figure 3.5.

### 3.3.3 Calculation of Privacy Score

By considering the $\beta_i$ as the sensitivity of each attribute and $F_{vis}(x)$ as the visibility of each attribute, the overall privacy disclosure score of each user can be calculated by privacy disclosure score function given by:

$$Privacy = \frac{\sum_{i=1}^{m} \beta_i * F_{vis}(xi)}{m}$$

where $i$ indicates the i-th attribute of a user and $m$ is the number of attributes. As the calculated value increases, it indicates that a user is more likely to have a risk of privacy and information disclosure, where less disclosure would be better.

## 3.4 Experimental Evaluation

In this section, the evaluation of the proposed privacy model is presented. Firstly the evaluation of the validation of the model is presented. Then, the accuracy of the model using real case studies is evaluated by comparing it with the privacy scoring model proposed by Liu and Terzi (2010a) which is the most recent polytomous approach for structured data.

### 3.4.1 Model Validation

To validate the privacy model, a simple test was undertaken in which three different cases were chosen: a user who has all the data public, a user who has all data private and a user who has some data public and some private. In a valid model, the obtained privacy scores should match the expected theoretical scores for the three chosen cases, that is, highest score, lowest score and in between, respectively. Table 3.4 illustrates the visibility scores obtained from the~~our~~ proposed model for three different cases. Clearly, the scores are fairly close to the theoretical expectation. For example, for user a, the visibility score is close to the highest expected score '8'. As in the proposed privacy function, sensitivity values are constant for a given user, therefore the privacy function validity is verified by the validity of the proposed visibility function.

Table 3.4: Users visibility comparison

| User | Inputs | | | Output | Expected Theoretical Score |
|---|---|---|---|---|---|
| | $F_{acc}$ | $F_{dif}$ | $F_{rel}$ | $F_{vis}$ | $F_{vis}$ |
| User a (All public data) | 4 | 0 | 1 | 7.32 | 8 |
| User b (All private data) | 0 | 3 | 0 | 0.884 | 0 |
| User c (Partially public data) | 2 | 1.5 | 0.5 | 4.12 | (3-5) |

### 3.4.2 Case Study

In the previous section, the author measured to what extent the users' personal information is revealed in multiple sources of online social networks. In other words, the intent is to calculate the privacy risk based on how much information a user has disclosed overall, in all the social networks. Here, the data were gathered from 15 users who were involved in four different online social networks (Facebook, ResearchGate, LinkedIn, and Google+) containing 11 attributes (as in Table 3.1) for each user, in order to measure the information disclosure and privacy risk of those users. The chosen number of users covers a diverse range of values from user profiles that is required to show the effectiveness of the proposed privacy score method. Then, to calculate the privacy disclosure score of users, two factors (sensitivity and visibility) that have a direct influence on users' privacy were considered. The sensitivity values were obtained from the literature review and historical work. In order to compute the visibility score, a Mamdani fuzzy inference system to obtain the visibility score after calculating the related functions (Accessibility, Difficulty of data extraction and Reliability) was deployed. Then,

the overall privacy disclosure score for each user was computed. At the final step, the privacy disclosure score of all the users with the privacy scoring model of Liu and Terzi (2010a) was compared to evaluate the accuracy of the proposed model.

Table 3.5: Two Users' Accessibility

| User (Accessibility) | User o (FB, RG, LD, G+) | User b (FB, RG, LD, G+) |
|---|---|---|
| Contact number | (2,2,4,3)→ $F_{acc}$ = 2.75 | (1,1,1,2)→ $F_{acc}$ = 1.25 |
| Email | (2,2,3,4)→ $F_{acc}$ = 2.75 | (1,1,1,2)→ $F_{acc}$ = 1.25 |
| Address | (2,2,2,4)→ $F_{acc}$ = 2.5 | (2,1,2,1)→ $F_{acc}$ = 1.5 |
| Birthdate | (3,2,3,4)→ $F_{acc}$ = 3 | (1,1,1,1)→ $F_{acc}$ = 1 |
| Home town | (3,2,3,4)→ $F_{acc}$ = 3 | (2,1,1,1)→ $F_{acc}$ = 1.25 |
| Current town | (3,3,2,4)→ $F_{acc}$ = 3 | (3,1,2,1)→ $F_{acc}$ = 1.75 |
| Job details | (2,4,4,4)→ $F_{acc}$ = 3.5 | (2,1,4,1)→ $F_{acc}$ = 3 |
| Relationship Status | (3,2,2,4)→ $F_{acc}$ = 2.75 | (2,1,1,1)→ $F_{acc}$ = 1.25 |
| Interests | (3,3,2,3)→ $F_{acc}$ = 2.75 | (2,1,3,1)→ $F_{acc}$ = 1.75 |
| Religious views | (3,2,2,4)→ $F_{acc}$ =2.75 | (1,1,1,1)→ $F_{acc}$ = 1 |
| Political views | (2,2,1,1)→ $F_{acc}$ = 1.5 | (1,1,1,1)→ $F_{acc}$ = 1 |

**Analysis of Results**

For the experiments, the value of $F_{acc}$ and $F_{dif}$ functions were calculated by considering the accessibility and difficulty values of each attribute for each user as the input. These values may vary in each social network.

By considering the user accessibility (Table 3.5) and difficulty of data extraction (Table 3.6), the final calculated values for the accessibility and difficulty score for two users (user o and user b - FB=Facebook, RG=ResearcGate, LD=LinkedIn, G+=Google+) are provided as a sample.

Table 3.7 illustrates the gained values from the fuzzy inference system. It should be noted that if the accessibility of all sources has the value equal to 1

Table 3.6: Two Users' Data Extraction Difficulty

| User (Difficulty) | User o (FB, RG, LD, G+) | User b (FB, RG, LD, G+) |
|---|---|---|
| Contact number | (2,2,3,3)$\rightarrow F_{dif}$ = 2.5 | (1,1,1,2)$\rightarrow F_{dif}$ = 1.25 |
| Email | (2,2,3,4)$\rightarrow F_{dif}$ = 2.5 | (1,1,1,2)$\rightarrow F_{dif}$ = 1.25 |
| Address | (2,2,2,3)$\rightarrow F_{dif}$ = 2.25 | (2,1,2,1)$\rightarrow F_{dif}$ = 1. 5 |
| Birthdate | (3,2,3,3)$\rightarrow F_{dif}$ = 2.75 | (1,1,1,1)$\rightarrow F_{dif}$ = 1 |
| Home town | (3,2,3,3)$\rightarrow F_{dif}$ = 2.75 | (2,1,1,1)$\rightarrow F_{dif}$ = 1.25 |
| Current town | (3,2,3,3)$\rightarrow F_{dif}$ = 2.75 | (3,1,2,1)$\rightarrow F_{dif}$ = 1.75 |
| Job details | (2,3,3,3)$\rightarrow F_{dif}$ = 2.75 | (2,1,3,1)$\rightarrow F_{dif}$ = 1.75 |
| Relationship Status | (3,2,2,3)$\rightarrow F_{dif}$ = 2.5 | (2,1,1,1)$\rightarrow F_{dif}$ = 1.25 |
| Interests | (3,3,2,3)$\rightarrow F_{dif}$ = 2.75 | (2,1,3,1)$\rightarrow F_{dif}$ = 1.75 |
| Religious views | (3,2,2,3)$\rightarrow F_{dif}$ = 2.5 | (1,1,1,1)$\rightarrow F_{dif}$ = 1 |
| Political views | (2,2,1,1)$\rightarrow F_{dif}$ = 1.5 | (1,1,1,1)$\rightarrow F_{dif}$ = 1 |

(only accessible to the user), the reliability of the data would be zero and that parameter is exempted from the final calculation for visibility (in other words, the visibility for that parameter is zero).

For the moment, the fuzzy inference system is used to calculate the visibility of each attribute for the users. Table 3.8 shows the obtained values for visibility.

By comparing the results of the table, it can be seen that users do not tend to provide the information, which is more sensitive than the other attributes. Based on the experiments, it was found that users are likely to disclose their information such as email addresses, current towns and interests, while information related to their political and religious views has less likelihood of disclosure.

After computing the visibility score of each attribute for the users, the next step is to calculate the overall privacy disclosure score for the users. Regarding the case study, which involves 15 users, a privacy disclosure score calculation was deployed, having been derived from the previous section.

Table 3.7: FIS-based visibility calculation

| User (Difficulty) | User o | | | | User b | | | |
|---|---|---|---|---|---|---|---|---|
| | $F_{acc}$ | $F_{dif}$ | $F_{rel}$ | $F_{vis}$ | $F_{acc}$ | $F_{dif}$ | $F_{rel}$ | $F_{vis}$ |
| Contact number | 2.75 | 2.5 | 0.96 | 7.32 | 1.25 | 1.25 | 0.46 | 1.5 |
| Email | 2.75 | 2.5 | 0.96 | 7.32 | 1.25 | 1.25 | 0.48 | 1.52 |
| Address | 2.5 | 2.75 | 0.97 | 7.3 | 1.5 | 1.5 | 0.76 | 4.11 |
| Birthdate | 3.2 | 2.75 | 0.96 | 7.32 | 1 | 1 | 0 | 0 |
| Home town | 3.2 | 2.75 | 0.96 | 7.32 | 1.25 | 1.25 | 0.46 | 1.5 |
| Current town | 3.2 | 2.75 | 0.96 | 7.32 | 1.75 | 1.75 | 0.76 | 4.11 |
| Job details | 3.5 | 2.75 | 0.96 | 7.32 | 3 | 1.75 | 0.76 | 7.32 |
| Relationship Status | 2.75 | 2.5 | 0.96 | 7.32 | 1.25 | 1.25 | 0.46 | 1.5 |
| Interests | 2.75 | 2.5 | 0.96 | 7.32 | 1.75 | 1.75 | 0.76 | 4.11 |
| Religious views | 2.75 | 2.5 | 0.96 | 7.32 | 1 | 1 | 0 | 0 |
| Political views | 1.5 | 1.5 | 0.76 | 4.11 | 1 | 1 | 0 | 0 |

Table 3.9 shows the computed results for the users and illustrates the final privacy disclosure score of the users in the case study. Regarding the obtained value for the privacy disclosure score of each user, it can be observed that the users who have greater willingness to disclose their information (both sensitive and non-sensitive) have higher risk for their privacy.

Figures 3.6 and 3.7 illustrate the overall privacy disclosure score of each user as measured by two different methods after data normalisation (between 0 and 1). In figure 3.6, the comparison of the results indicates that majority of calculations applying FIS model exhibit higher disclosure scores, excluding two exceptions of 'i' and 'l'. This generally higher level of privacy disclosure score from FIS model is because of using a higher number of input data from multiple sources which results in revealing more information, as well as the accuracy and reliability of each attribute itself.

For better clarification, three users with three different patterns have been

Table 3.8: FIS model for visibility score calculation

| Users | Contact Number | Email | Address | Date of Birth | Home Town | Current Town | Job Details | Relationship Status | Interests | Religious Views | Political Views |
|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 3.59 | 3.6 | 3.6 | 3.5 | 1.5 | 2.65 | 3.59 | 0 | 0 | 0 | 0 |
| b | 1.51 | 1.52 | 4.11 | 0 | 1.5 | 4.1 | 7.32 | 1.5 | 4.1 | 0 | 0 |
| c | 4.1 | 3.6 | 1.01 | 1.64 | 1.01 | 5.66 | 5.26 | 0 | 4.12 | 3.54 | 0 |
| d | 1.52 | 1.5 | 1.5 | 1.5 | 0 | 4.1 | 4.12 | 0 | 4.12 | 2.64 | 4.14 |
| e | 1.5 | 7.33 | 4.42 | 4.1 | 4.12 | 7.32 | 7.32 | 0 | 4.12 | 0 | 0 |
| f | 1.5 | 1.51 | 4.11 | 4.1 | 0 | 7.32 | 7.32 | 4.1 | 4.12 | 4.12 | 0 |
| g | 4.12 | 4.12 | 5.66 | 4.11 | 1.5 | 5.66 | 5.66 | 1.5 | 7.32 | 4.11 | 0 |
| h | 2.65 | 2.65 | 7.32 | 4.12 | 0.88 | 7.16 | 2.65 | 2.59 | 7.31 | 2.64 | 1.01 |
| i | 4.1 | 7.3 | 7.32 | 4.11 | 4.12 | 4.12 | 7.31 | 1.51 | 7.3 | 1.01 | 1.51 |
| j | 7.32 | 4.12 | 4.41 | 4.12 | 4.11 | 7.31 | 7.32 | 1.5 | 4.1 | 4.11 | 1.5 |
| k | 4.1 | 4.12 | 4.1 | 4.1 | 4.11 | 7.3 | 7.32 | 4.11 | 7.32 | 4.12 | 4.12 |
| l | 3.59 | 2.64 | 7.32 | 0.88 | 4.12 | 7.32 | 7.32 | 1.2 | 7.3 | 7.3 | 1.01 |
| m | 4.11 | 4.11 | 4.12 | 3.54 | 7.32 | 7.3 | 7.3 | 0 | 7.32 | 7.3 | 4.11 |
| n | 4.1 | 1.52 | 3.54 | 4.11 | 4.12 | 7.32 | 7.3 | 4.11 | 7.18 | 4.12 | 7.3 |
| o | 7.32 | 7.3 | 7.32 | 7.32 | 7.3 | 7.31 | 7.32 | 7.32 | 7.32 | 7.31 | 4.11 |

Table 3.9: User's final privacy disclosure score calculation

| Users | Contact Number | Email | Address | Date of Birth | Home Town | Current Town | Job Details | Relationship Status | Interests | Religious Views | Political Views | Privacy Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 2.154 | 0.659 | 1.060 | 0.408 | 0.225 | 0.308 | 0.718 | 0 | 0 | 0 | 0 | 0.684 |
| b | 0.904 | 0.278 | 3.493 | 0 | 0.225 | 0.478 | 1.464 | 0.624 | 1.230 | 0 | 0 | 0.790 |
| c | 2.460 | 0.659 | 0.858 | 0.191 | 0.151 | 0.599 | 1.252 | 0 | 1.215 | 2.009 | 0 | 0.861 |
| d | 0.912 | 0.273 | 1.275 | 0.174 | 0 | 0.478 | 0.824 | 0 | 1.218 | 1.493 | 2.828 | 0.863 |
| e | 0.900 | 1.341 | 3.757 | 0.478 | 0.618 | 0.851 | 1.464 | 0 | 1.236 | 0 | 0 | 0.967 |
| f | 0.900 | 0.276 | 3.493 | 0.478 | 0 | 0.853 | 1.464 | 1.708 | 1.215 | 2.334 | 0 | 1.158 |
| g | 2.472 | 0.754 | 4.811 | 0.479 | 0.225 | 0.659 | 1.132 | 0.624 | 2.195 | 2.328 | 0 | 1.429 |
| h | 1.590 | 0.485 | 5.222 | 0.480 | 0.132 | 0.834 | 0.530 | 1.078 | 2.193 | 1.495 | 0.690 | 1.430 |
| i | 2.460 | 1.336 | 5.222 | 0.479 | 0.618 | 0.480 | 1.462 | 0.629 | 2.190 | 0.372 | 1.031 | 1.589 |
| j | 4.392 | 0.754 | 3.748 | 0.480 | 0.616 | 0.852 | 1.464 | 0.624 | 1.230 | 2.328 | 1.031 | 1.593 |
| k | 2.460 | 0.754 | 3.485 | 0.478 | 0.615 | 0.851 | 1.464 | 1.712 | 2.196 | 2.334 | 2.813 | 1.742 |
| l | 2.154 | 0.483 | 5.222 | 0.102 | 0.618 | 0.851 | 1.464 | 0.499 | 2.190 | 4.136 | 0.690 | 1.764 |
| m | 2.466 | 0.752 | 3.502 | 0.412 | 1.095 | 0.851 | 1.460 | 0 | 2.196 | 4.136 | 2.808 | 1.798 |
| n | 2.460 | 0.278 | 3.009 | 0.479 | 0.618 | 0.863 | 1.460 | 1.712 | 2.154 | 2.334 | 4.988 | 1.849 |
| o | 4.392 | 1.336 | 5.222 | 0.853 | 1.095 | 0.852 | 1.464 | 3.049 | 2.195 | 4.141 | 2.808 | 2.582 |

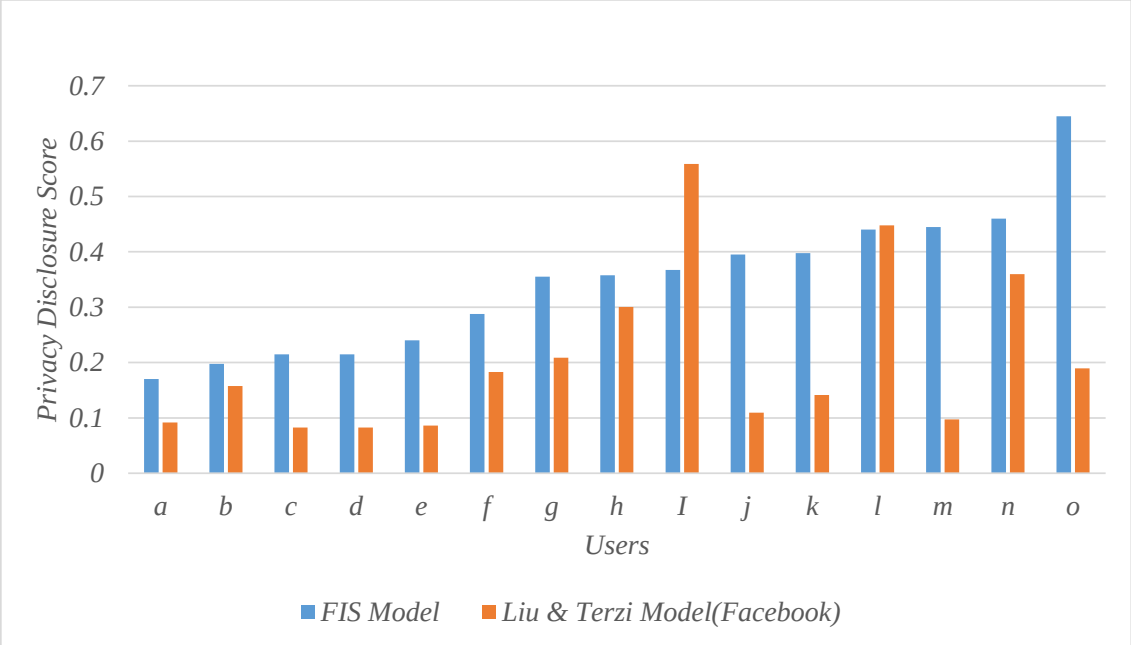



Figure 3.6: Comparison of FIS and Liu model (Facebook case) for PDS calculation

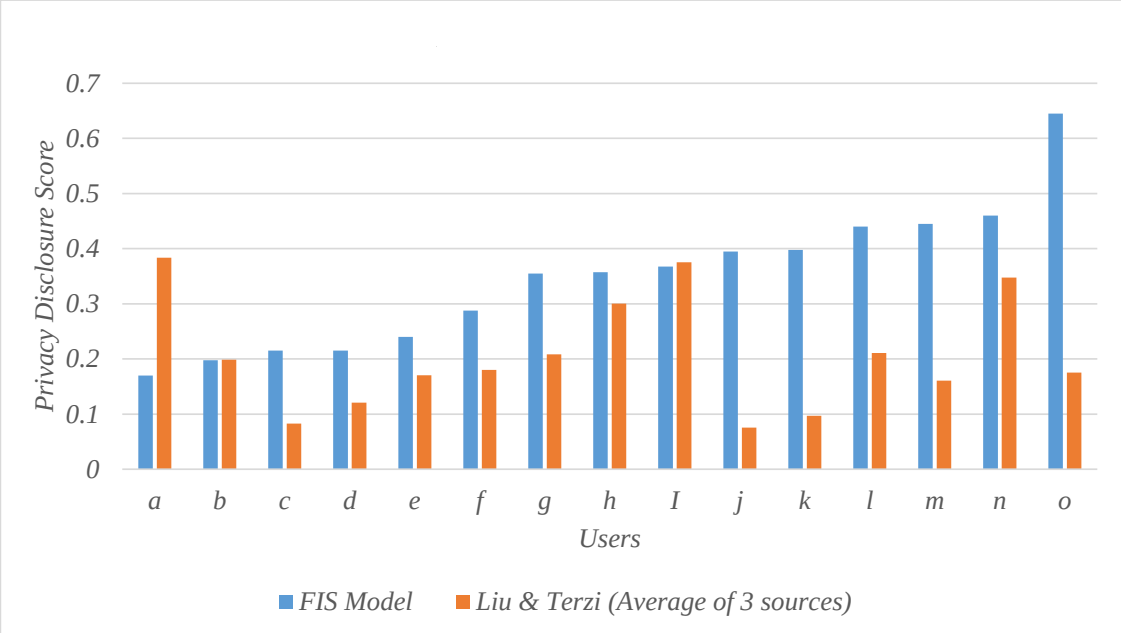



Figure 3.7: Comparison of FIS and Liu model (Average of 3 sources) for PDS calculation

compared. For the first case, the user 'i' exhibits an exceptionally higher privacy disclosure score using the Liu & Terzi model (0.56 for the Liu & Terzi (Liu and Terzi, 2010a) model vs 0.37 for FIS model) as a reason for sharing a high level of sensitive information within one source of input data. This user has shared most information on Facebook and did not provide a sufficient amount of sensitive information on other sources resulting in a lower level of reliability of data (which is calculated in the fuzzy phase) and consequently a decrease in the privacy disclosure score calculated by the FIS model. Hence, the obtained value by the Liu & Terzi method is higher which is reasonably practical for a single source of data. For the second case, user 'l', similarly indicates comparable disclosure scores for both models ( 0.45) due to the high proportion of released data in Facebook as one of the sources of information. Comparison of results obtained from users 'l' and 'o' indicates that both users have revealed the same amount of information on their Facebook profiles (6 out of 11 attributes). For the last case, user 'l' has disclosed more sensitive information than user 'o', resulting in a higher exposure score for user 'l' with the Liu & Terzi model. While the majority of the information shared by user 'o' on Facebook is not sensitive, other sources provide publicly available sensitive information. Therefore, unlike the obtained low level of disclosure score using the Liu method, the privacy disclosure score for user 'o' computed by the fuzzy method is very high.

Figure 3.7 illustrates the results from the FIS model with the average score of three social network sources (i.e. Facebook, LinkedIn and ResearchGate) calculated using the Liu & Terzi model. The graph clearly shows that the obtained average score is not sufficient to capture the risk of disclosure across multiple sites, excluding the user 'a'. User 'a' exhibits a significantly lower score of privacy disclosure by FIS model because of the scattered distribution of the attributes within each individual source. In this case, the reliability of data in the fuzzy function may not provide a high value as does the Liu & Terzi model.

## 3.5 Comparison with Other Techniques

In measuring privacy in online social networks, it is not inherently clear which information can result in a significant loss such as identity theft. Other risks are even harder to measure; comments about and pictures of a user, which are risk-free for some individuals, can be detrimental to others. One likely case is a criticism against a religion or government. In some countries and cultures, such criticism is broadly accepted whereas, in other countries, an individual can get in severe difficulties for making such comments (Renner, 2010; Bonti et al., 2012). Another risk of using online social networks is posting vacation information when users are abroad, intruders decide when to rob the house based on the information they gather.

As discussed in Chapter 2, Section 2.3.3, several techniques and methods have been proposed to calculate and compute privacy and information sharing in a public manner, including algorithmic and statistical approaches (Maximilien et al., 2009; Renner, 2010; Becker, 2009; Srivastava and Geethakumari, 2013; Liu and Terzi, 2010a; Anderson, 2013) which state the significance of quantifying privacy in an online social network. This issue becomes even more critical in the case of protecting the huge volume of corresponding personal information, especially in large-scale online social networks. Table 3.10 shows the comparison of previous key methods for measuring and calculating the privacy of online social networks and public data. All these studies consider measuring privacy risks and information leakage for the user from only a single source of data. Beside the aforementioned studies, several authors proposed tools for the configuration of privacy settings in specific online social networks. As shown in Table 3.10, previous methods had considered only a single source of online social networking sites, while in the proposed model, multiple online social networking sites are considered to calculate the privacy disclosure score of the users. Moreover, except for the Liu and Terzi (2010b) method, all other methods and models applied

the dichotomous approach (data is publicly available or private) for computing the privacy score, while their proposed model is the only polytomous-based one. The other point of uniqueness in the proposed model is its independence from data-type and data structure. In other words, as the level of visibility and the sensitivity of information which is shared in online social media is considered, the proposed model is able to measure the privacy risk for that user. Hence, based on information extracted from a social networking site, the proposed model can be run and applied to that data.

## 3.6 Summary

In this chapter two aspects were considered to compute the overall privacy disclosure score of a user who is participating in multiple social networks, namely, sensitivity and visibility. The information sensitivity was obtained from prior studies for use in the proposed system. Next, information visibility was computed as it has a direct impact on a user's privacy disclosure score by applying the fuzzy technique. Regarding the obtained privacy disclosure scores, it is concluded that users' privacy disclosure scores depend directly on the amount of information a user discloses, such as religious views, political views and relationship status. Also, the results obtained in this study allow the conclusion that considering information from multiple social sites gives a more accurate estimation of privacy risk. The work is formulated as a working formula that calculates the privacy disclosure score of each user while user's data is gathered from multiple sources of online social networks (such as Facebook, Twitter, LinkedIn, ResearchGate, and Google+). There is no limitation of the solution to a specific social network site but the proposed model can be deployed in all social networks. Finally, users can be informed about their privacy levels and how much data they have shared in such networks. In the next chapter, it will be shown how the privacy mea-

Table 3.10: Comparison of historical studies

| Reference | Focus | No. of Sources | Data Type | Approach |
|---|---|---|---|---|
| Liu and Terzi (2010a) | Privacy risk from individual perspective | 1 | Structured | Dichotomous & Polytomous |
| Srivastava and Geethakumari (2013) | Privacy risk of text messages | 1 | Unstructured | Dichotomous |
| Domingo-Ferrer et al. (2008) | Obtained utility by sharing information | 1 | Structured | Dichotomous |
| Nepali and Wang (2013) | Privacy exposure based on known parameters | - | Web data | Dichotomous |
| Becker (2009) | Attribute inference | 1 | Structured | Polytomous |
| Talukder et al. (2010) | Sensitive information leakage of a profile | 1 | Structured | Dichotomous |

surement can be undertaken for shared unstructured data within multiple online social media sites.

# Chapter 4

# An Automated Model to Score the Privacy of Unstructured Information

*"People get a little sidelined thinking that fame and fortune is going to bring them happiness, peace and contentment in their lives. Everyone thinks they want to be famous until the paparazzi are in their face, and then they're asking, Just give me some privacy." (Linda Thompson)*

Measuring the privacy of unstructured data caused from textual information comes with difficulties as it is not clear what metrics are influencing the privacy of the sentiments. Although there are various studies of privacy evaluation from the structured information extracted from unstructured data, there are limited privacy scoring methods concentrating on the views of the individuals and these cannot appropriately provide an accurate privacy score of shared unstructured data in social networks. Here, in this chapter, an automated fuzzy based model is proposed that can extract the privacy-related features, as well as the related shared structured data, and measure and warn users regarding the level of privacy risk they have on online social platforms. The evaluation of the proposed model indicates that it can measure users' privacy risks in a more accurate manner compared with previously proposed methods and available commercial software in the domain.

## 4.1 Introduction

As mentioned in the previous chapter, it not easy to estimate privacy as the privacy settings are time-consuming and complicated. Moreover, many of the online social network users may not be competent in estimating the privacy risk (Aghasian et al., 2017), and when it comes to content sharing, an approximation of privacy risk becomes harder as there are insufficient privacy settings related to unstructured data (Fiesler et al., 2017). Several comments about a user may exist in his/her friends' social profiles and this is not easy to control. Hence, there is a need for mechanisms to measure or preserve the privacy of the users who share information on social media. It is not inherently clear what factors can influence the privacy calculation of the unstructured data. As the intention of the shared sentiment may vary depending on context and terms can have a different meaning in a shared sentiment, it is not easy to handle unstructured data in terms of privacy calculation.

In this regard, to measure the privacy risk, there are a few methods proposed to score the information leakage in social networking sites for unstructured data. Farzindar and Inkpen (2015); Wilson et al. (2005) have focused on the polarity of the sentiments regardless of the risk score for the shared information. This is also known as orientation which is the emotion expressed in the sentence. It can be positive, negative or neutral. Polarity in sentiment analysis refers to identifying sentiment orientation [positive, neutral, and negative] in written or spoken language. Other researchers (Khazaei et al., 2018; Canfora et al., 2018) focused on the polarity of the sentiments or the number of false positives (a false positive is an outcome where the model incorrectly predicts the positive class) and true negatives (is an outcome where the model correctly predicts the negative class) of unstructured data rather than the exact amount of privacy risk caused by the shared sentiments. As users share different free-format texts in multiple online social sites, more information about them can be mined. Hence, the privacy risk

may increase. Accordingly, to provide an accurate privacy scoring model, there is a need to consider overall shared information including both structured and unstructured across all of the social media platforms in which a user participates.

The aim of this chapter is to explore the causes of privacy risks for users of social media sites and to propose a privacy scoring model in social media to inform users about the level of their privacy in the unstructured information they have shared. Moreover, the integration of the proposed approach to social media platforms could help users to be informed of the risk of what they are writing before they share it.

The rest of this chapter is organised as follows: Section 4.2 explains the problem definition and provides the overall design of the model for measuring the privacy score of shared sentiments. Section 4.3 presents the evaluation of the proposed model applying real social network data. Section 4.4 discusses the comparison with other techniques in text mining and proposed privacy scoring models. Section 4.5 addresses the conclusion of the study.

## 4.2 Problem Definition & Privacy Scoring Model

As privacy breaches and risks for online social media users increase, there is a need to provide protective methods to ensure safety on such social media sites or to control users' personal information and enable them to perform their online activities in a safe manner (Bell, 2014; Lipschultz, 2014). One way to tackle the users' safety issues on online social networks is to measure the privacy risk. Measuring the users' privacy for their shared information on online social networking sites is a challenging task. It is also becoming more challenging when it comes to unstructured data as a term may be sensitive in a post or comment while it does not influence the privacy of the user in another tweet, post or comment.

This study considers the following questions:

- What pieces of information in a context may lead to privacy breach or risks for users on online social networks?

- How should the views of users on what causes privacy risks for them on online social networking sites be measured?

In this case, one important factor that should be considered to score the privacy is to comprehend the views of individuals. This becomes more vital when there are multiple data sources and users disclose their information on various social sites. As more information in the form of unstructured data is being shared, the probability of privacy risks increases. Since users disclose both structured and unstructured information on social networking sites, one cannot rely solely on unstructured data to score the privacy of users or decide the polarity of the sentiment. The background knowledge of users, which is shared among social network sites, is an essential factor that should also be brought into consideration. For example, imagine the following sentiment shared by two different users (User $a$ – from country $x$, User $b$ – from country $y$):

$$Sentimentsample = \begin{cases} \text{user } a \text{ \& I hate country '}x\text{'} \\ \text{user } b \text{ \& I hate country '}x\text{'} \end{cases}$$

As can be observed, both users have shared identical statements. In this case, depending on the background information such as religious belief and country they are from, same statement can be interpreted with different sentiments. The shared text by user $a$ may not be sensitive while the statement shared by user $b$ may include racist content which is a negative polarity. Hence, besides the text itself, there is a need to know some background details about the users, such as age range, country, religion or political views, if applicable, in order to provide a good estimation of privacy risk for shared statements.

Figure 4.1 shows the overall flow of the privacy scoring process of the pro-

posed model. The model is comprised of three different phases. In the first phase, information retrieval and pre-processing of the data occur. Then, the privacy score of the sentiment is measured by considering some background information of users based on fuzzy systems. Finally, the output of the proposed system provides the privacy score of users who have shared information on online social networking sites.

In the proposed model, unstructured data, which will be gathered from online social networks including tweets, comments and posts, is used to score privacy. Sensitive information or personality traits of users which have been shared through these social network sites are also considered in scoring privacy. Moreover, some background information of users will be collected to justify the credibility of the model. Here it is presumed that each user participates in multiple online social networks and different types of free-format texts are revealed to other users, depending on the nature of the social media. Unlike other methods, the aims are to see what can be intercepted from shared unstructured data, to score privacy and to warn users based on the level of vulnerability and importance of that piece of shared information.

### 4.2.1 Information Retrieval and Pre-processing

Before processing the sentiments of the users, there is a need to understand what features of a sentiment may lead to a privacy breach in the shared textual information of users of online social media. To achieve this, an open dataset of social networks from *Kaggle*[1] repository for the training set was acquired. The dataset contains the tweets, posts and comments of public figures and also incorporates different information about sport, politics, health and other domains. The aim of the message has been specified in the dataset as well.

[1]https://www.kaggle.com/

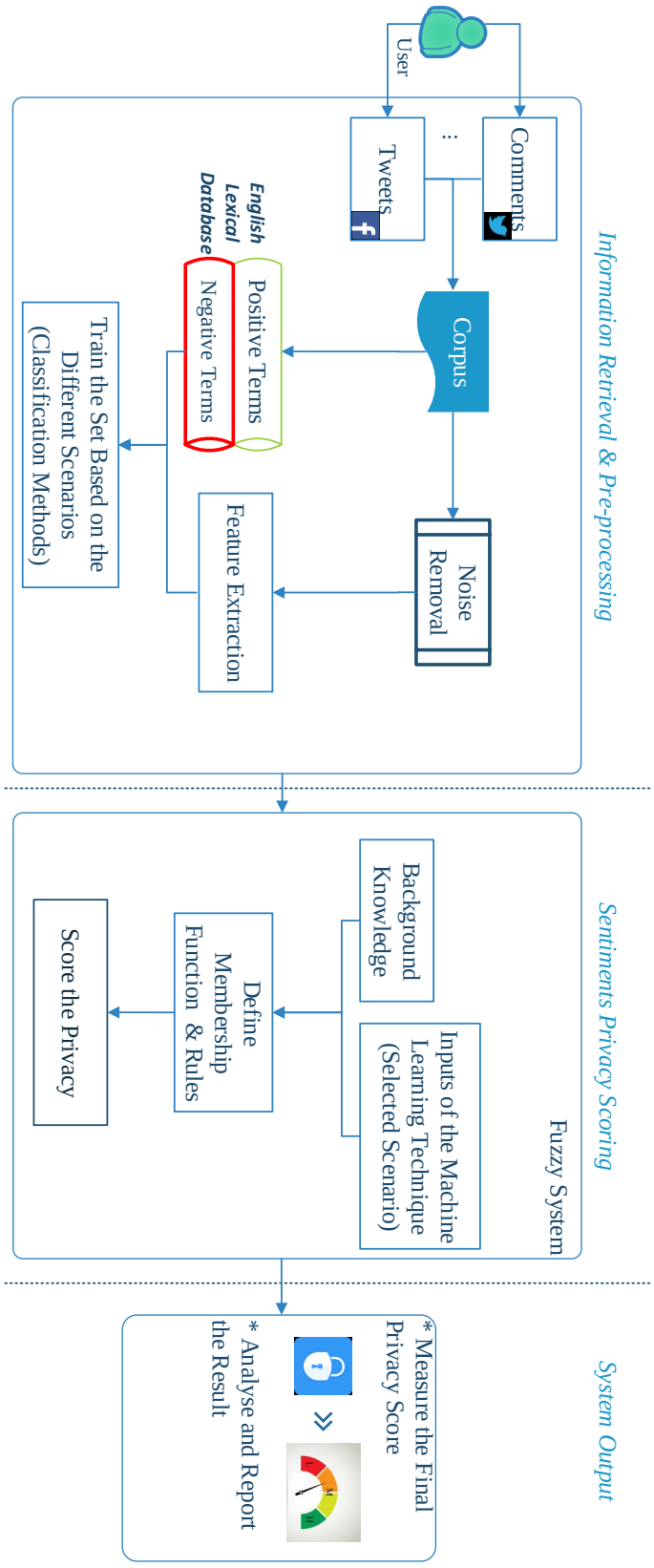



Figure 4.1: Flow of Privacy Scoring Process of Proposed Method

To process the unstructured data, firstly, stop-words and the noise are removed. Stop-words include any words with two or fewer letters which exist in the document and also some words which do not affect the meaning of the unstructured data like 'the'. These words have been specified in a separate dictionary. Next, the noise is removed from the documents. Noise includes any sign or marks (like exclamation marks or question marks).

To extract the features, firstly, both uni-grams and POS bi-grams tagging techniques (Toutanova et al., 2003) were applied to extract meaningful terms from the unstructured data of each user. By mapping the extracted terms with the index dictionary,[2] it is possible to find out which terms are positive, negative or neutral in a context. The phrases can contain nouns, verbs or adjectives which are extracted from *Wordnet* English lexical database (Miller, 1995).

In the next step, to train the model, three different scenarios have been considered to find out which one would be best fitted into the model. The considered scenarios are as follow:

- Only the shared unstructured data is considered to train the model.
- Only the number of positive and negative phrases in each sentiment is considered.
- The model is trained by accumulating the unstructured data and the number of positive or negative words.

In the first scenario, the features that have an impact on the privacy of users such as negation, adjectives and adverbs regarding their F-Score, are considered. In the second scenario, for understanding the informative phrases within a text, the tfidf numerical statistic has been applied to consider the importance of a word in a sentiment. Considering $d$ as the document and $t$ as the term, the frequency of the term is calculated by (Schütze et al., 2008):

[2]The index dictionary contains three different dictionaries including positive terms, negative terms and neutral.

$$\mathrm{tf}(\mathrm{t},\mathrm{d}) = log(1 + \mathrm{f}_{\mathrm{t,d}}) \tag{4.1}$$

if $\mathrm{f}_{\mathrm{t,d}}$ is higher than 0. Otherwise, the result is equal to zero. The reason for applying the logarithmic formula is related to the probability of the occurrence of a term in a document. (It is not likely that $x$ occurrences of a phrase in a document actually carry $x$ times the importance of a single occurrence) Next, the inverse document frequency is calculated by:

$$\mathrm{idf}(\mathrm{t},\mathrm{D}) = log\frac{N}{|\{d \in D : t \in d\}|} \tag{4.2}$$

where $N$ is the sum of documents in corpus. At last, the tfidf is calculated by:

$$\mathrm{tfidf}(\mathrm{t},\mathrm{d},\mathrm{D}) = \mathrm{tf}(\mathrm{t},\mathrm{d}) \cdot \mathrm{idf}(\mathrm{t},\mathrm{D}) \tag{4.3}$$

In the third scenario, model training is achieved by accumulating the unstructured data and the number of positive or negative words. The comparison of the model accuracy of the applied supervised machine learning methods (both Naive Bayes and J48 decision trees), considering cross-validation, enables the choice of the most appropriate technique. Then the rules for scoring the privacy are defined. It should be noted that other machine learning classification mechanisms have been applied for the hypothesised scenarios and Naive Bayes and J48 trees have provided comparatively better results than other techniques.

In the test phase, a corpus has been created which contains any shared information by users which can be in the form of tweets, posts or the comments containing some background information. After selecting the best fitting scenario, the gained information, as an input to define the membership function of the fuzzy system, was applied. Also, the background knowledge is considered as another input for privacy scoring. By applying the fuzzy system, it can be decided whether the risk of shared unstructured data is low, medium or high.

Table 4.1: Membership Function Database

| LV | Type | MF | Range | a | b | c | d | Applied MF |
|---|---|---|---|---|---|---|---|---|
| No. of Features | Input | Low | [0,10] | 2.82 | 3 | - | - | Spline-based |
| No. of Features | Input | Medium | [0,10] | 109 | 294 | 61.7 | 6.04 | Sigmoid Derivative |
| No. of Features | Input | High | [0,10] | 24.2 | 6.02 | - | - | Sigmoid |
| Positive Terms | Input | Low | [0,8] | 1.9 | 2.01 | - | - | Spline-based |
| Positive Terms | Input | Medium | [0,8] | 300 | 1.98 | 163 | 4.98 | Sigmoid Derivative |
| Positive Terms | Input | High | [0,8] | 3.1 | 5 | - | - | Sigmoid |
| Negative Terms | Input | Low | [0,8] | 0.98 | 1.01 | - | - | Spline-based |
| Negative Terms | Input | Medium | [0,8] | 61.9 | 1.01 | 1.48 | 2.96 | Sigmoid Derivative |
| Negative Terms | Input | High | [0,8] | 2.95 | 2.98 | - | - | Sigmoid |
| Background Knowledge | Input | Low | [0,1] | 0.1 | 0.199 | - | - | Spline-based |
| Background Knowledge | Input | High | [0,1] | 0.153 | 0.617 | - | - | Sigmoid |
| Privacy Score | Output | Low | [0,1] | 0.242 | 0.31 | - | - | Spline-based |
| Privacy Score | Output | Medium | [0,1] | 325 | 0.31 | 370 | 0.6 | Sigmoid Derivative |
| Privacy Score | Output | High | [0,1] | 142 | 0.62 | - | - | Sigmoid |

Linguistic Variable (LV), Membership Function (MF)

### 4.2.2 Unstructured Data Privacy Score

By extracting the features and mapping it to the obtained results from the training set and accumulation with the background knowledge, the rules and membership functions for the fuzzy system are defined. By defining the membership function, the related bound of privacy for each tweet, post or comment can be obtained which provides the final privacy score of that piece of information for a user and gives support to cluster the level of risk by low, medium and high. Based on the frequency of the informative terms (existing negative and positive terms in a sentiment) and obtained results of classification, the membership function of the fuzzy system is defined. Table 4.1 illustrates the detailed membership functions and rules applied in the fuzzy system model.

Also, as discussed in Section 4.2, the background knowledge is considered as another privacy measure. Hence, in the fuzzy system, applying the background knowledge related to the shared unstructured data can increase the credibility of

the measurement and can provide a better understanding for users. The background knowledge of the users indicates if any structured information shared about a user exists in the same domain of the sentiment. This information can be obtained from either the same social media profile or other social media profiles in which the user participates. Hence, the functions are defined in a way that they provide the most precise results based on the input of the model. A user who has shared a sentiment regarding his employment situation can be considered. If the user shares any background information regarding his/her job in the structured section of the profile, then the level of sensitivity of that sentiment increases. Hence, the privacy risk for the user is heightened compared with a case in which the user has not shared any information in his/her social media profile.

Excluding the membership functions which have been defined in Table 4.1, a fuzzy rules database has been defined to create a logical calculation for the privacy of the users' shared sentiments based on the inputs (Table 4.2). For the aggregation of the rules, a *max* operation has been applied. The defined rules for the fuzzy system are obtained based on the machine learning results and consultation with the professionals in the domain knowledge.

As has been discussed in (Aghasian et al., 2017), the process of a fuzzy system contains four steps including fuzzification of inputs, implication, aggregation and defuzzification. To obtain a crisp number from the defined rules in the fuzzy system, the sigmoid function has been applied. The application of the sigmoid function and its derivatives provides the final output of the system based on the provided metrics obtained from the previous phase. As the sigmoid function is essentially open to the right, it can appropriately respond to the abstraction such as very high. The sigmoid function is given by:

$$S(x) = \frac{1}{1 + e^{-x}} \tag{4.4}$$

where $e$ is the Euler's number (natural logarithm) and $x$ indicates the value of

Table 4.2: Fuzzy Rules Database

| | Inputs | | | | Output |
|---|---|---|---|---|---|
| | No. of Features | No. of Positive Terms | No. of Negative Terms | Background Knowledge | |
| 1 | L | X | L | X | L |
| 2 | L | X | M | L | M |
| 3 | L | X | M | H | H |
| 4 | L | X | H | X | H |
| 5 | M | X | L | L | L |
| 6 | M | X | L | H | M |
| 7 | M | X | M | L | M |
| 8 | M | X | M | H | H |
| 9 | M | X | H | X | H |
| 10 | H | L | L | L | L |
| 11 | H | L | L | H | M |
| 12 | H | X | M | L | M |
| 13 | H | X | M | H | H |
| 14 | H | X | H | X | H |

L= Low, M=Medium, H=High, X=Can be in any state

midpoint of the sigmoid function.

It should be noted that the reason for applying the fuzzy system for calculating the privacy score is related to the uncertainties which exist in natural language and the grade of a system which is an interval rather than a crisp number. It has been proved that the fuzzy system is a scientifically accurate scheme for linguistic uncertainties modelling (Mendel and Wu, 2006).

### 4.2.3 System Output

After measuring the privacy score of the sentiments, the results were analysed and a report was provided for users to realise the level of privacy of that piece of information, and its polarity. By implementing this, the proposed method helps users to understand the level of privacy risk related to the shared sentiments belonging to them and to consider mitigation actions to reduce the negative impacts

of shared free-text data.

## 4.3 Experimental Evaluation

This section presents the evaluation of the proposed privacy scoring model. At first, the evaluation and accuracy of the machine learning techniques are presented. Then, the evaluation and comparison of the fuzzy system is illustrated by applying a real case study with a previously proposed privacy scoring model, and linguistic inquiry and word count (*LIWC*) software which is the most relevant one for text analysis and privacy computation. The proposed system is evaluated with real online social media data obtained from open datasets.

### 4.3.1 Machine Learning Methods

To select the inputs for the fuzzy system, three different scenarios were evaluated to decide which model can be best fitted into the proposed model to apply in the fuzzy system. To achieve this, a dataset containing 494 instances of tweets, comments and posts from both Facebook and Twitter online social networking sites has been considered as the input corpus (including 33 highly sensitive, 321 with medium sensitivity and 140 instances with low sensitivity). Firstly, the features as well as the number of negative and positive terms for each instance of the corpus are extracted and then compared the F-score and precision of the machine learning methods for different scenarios. Both Naive Bayes and J48 decision tree machine learning techniques with 10-fold cross-validation were utilised to determine which scenario comes with the best F-score and precision. Table 4.3 shows the obtained results of different scenarios by machine learning methods. Although the algorithms in the second scenario provide a better precision and F-score compared with the other scenarios, the confusion matrices do not pro-

Table 4.3: Comparison of Machine Learning Techniques

| Scenario | ML Method | Estimate | F-Score | Precision | Recall | 95% CI |
|---|---|---|---|---|---|---|
| Text Only | Naive Bayes | 0.59 | 0.4 | 0.41 | 0.38 | 0.56–0.62 |
| | J48 Decision Tree | 0.57 | 0.37 | 0.37 | 0.36 | 0.54–0.6 |
| P+N Terms | Naive Bayes | 0.64 | 0.51 | 0.42 | 0.64 | 0.6–0.68 |
| | J48 Decision Tree | 0.64 | 0.51 | 0.41 | 0.65 | 0.6–0.68 |
| Text+P+N | Naive Bayes | 0.61 | 0.42 | 0.43 | 0.40 | 0.58–0.64 |
| | J48 Decision Tree | 0.58 | 0.35 | 0.35 | 0.33 | 0.55–0.61 |

ML= Machine Learning, CI=Confidence Interval of Estimate, P=Positive, N=Negative

vide appropriate prediction for low and high classes. The obtained confusion matrices 4.4-4.6 lay out the occurred errors by the classification techniques and visualise the algorithm performance. The obtained confusion matrices can help to incur a better judgment about which scenario can better fit to calculate the privacy based on the fuzzy system. For the second scenario, the classifier can only label the data which comes with high density in the dataset. Unlike the confusion matrix for the second scenario (Table 4.5), the Naive Bayes algorithm of the third scenario (Table 4.6-a) provides a slightly better proportion of classification compared with other methods. Hence, the Naive Bayes algorithm from scenario 3 is chosen which comes with the highest precision. As the number of values belonging to each class is not balanced, stratified sampling may improve the precision of the machine learning methods. Stratification would help to provide the class proportion of each fold the same as the main dataset which can principally assure an impartial split between the test and train set. Stratification can preserve the distribution of class labels of the dataset in each fold that has been created by the cross-validation. In the proposed classification models, the stratification can improve the accuracy of the scenario by about 1 to 2 percent.

By comparing the obtained results from machine learning classification methods in three different scenarios, it cannot be seen that the classification algorithms of the second scenario have a better model accuracy compared with other tech-

| Act/Pred | High | Medium | Low |
|---|---|---|---|
| High | 2 | 30 | 1 |
| Medium | 7 | 254 | 60 |
| Low | 2 | 98 | 40 |

(a) Naive Bayes Technique

| Act/Pred | High | Medium | Low |
|---|---|---|---|
| High | 3 | 26 | 4 |
| Medium | 17 | 247 | 57 |
| Low | 5 | 102 | 33 |

(b) J48 Decision Tree Technique

Table 4.4: Confusion Matrices, Scenario 1

| Act/Pred | High | Medium | Low |
|---|---|---|---|
| High | 0 | 33 | 0 |
| Medium | 1 | 320 | 0 |
| Low | 1 | 139 | 0 |

(a) Naive Bayes Technique

| Act/Pred | High | Medium | Low |
|---|---|---|---|
| High | 0 | 33 | 0 |
| Medium | 0 | 321 | 0 |
| Low | 0 | 140 | 0 |

(b) J48 Decision Tree Technique

Table 4.5: Confusion Matrices, Scenario 2

| Act/Pred | High | Medium | Low |
|---|---|---|---|
| High | 2 | 30 | 1 |
| Medium | 9 | 254 | 58 |
| Low | 2 | 96 | 42 |

(a) Naive Bayes Technique

| Act/Pred | High | Medium | Low |
|---|---|---|---|
| High | 3 | 26 | 4 |
| Medium | 13 | 249 | 59 |
| Low | 4 | 102 | 34 |

(b) J48 Decision Tree Technique

Table 4.6: Confusion Matrices, Scenario 3

niques. However, based on the confusion matrix, the second scenario methods can only classify one class (the label which has a high number of instances related to that class). Meanwhile, the Naive Bayes technique in the third scenario has the best model accuracy after the second scenario and provides the best confusion matrix (fewest false negatives). In this phase, the number of features which have impact on the classifier were investigated and then the obtained result applied as the inputs for the fuzzy system.

It should be noted that based on the obtained results from the supervised machine learning algorithms, it can be comprehended that machine learning itself cannot provide an acceptable level of accuracy for prediction of the privacy level. Hence, the fuzzy system is applied to improve the model accuracy to score the privacy level of a shared unstructured data in online social networking site.

Table 4.7: Confusion Matrix, Fuzzy-based Proposed Model

| Proposed Model Overall Statistics | | Predicted | | |
|---|---|---|---|---|
| | | High | Low | Medium |
| Reference | High | 4 | 0 | 0 |
| | Low | 0 | 10 | 1 |
| | Medium | 6 | 3 | 26 |
| Balanced Accuracy | | 0.7 | 0.87 | 0.78 |

### 4.3.2 Fuzzy-based Privacy Risk Calculation

In this section, the privacy score of a case study containing 50 sentiments (one fold of the training set in the machine learning technique with no label) is examined. Regarding the defined fuzzy rules and obtained features from the machine learning model (as discussed in Section 4.2.2, the inputs of the fuzzy system have been obtained from the trained machine learning system), the unstructured data privacy risk is examined. After obtaining the result, a comparison of the result with a previously proposed model and a built-in software (*LIWC*) is undertaken for validation. Table 4.7 presents the confusion matrix and the balanced accuracy of the proposed model. As can be seen from the Table 4.7, considering the users' background information can improve the prediction accuracy of different classes although there are some errors. These errors pertain to the number of features of the sentiments; as the length is becoming shorter, the probability of the error increases. Furthermore, balanced accuracy provides the prediction accuracy for each class indicating the fraction of correctly predicted positives among all positives.

Moreover, Figure 4.2 illustrates the comparison of the proposed model and its 95% lower and upper confidence interval with a previously proposed model in online social networking sites.

The comparison of the test data and the previously proposed model (as depicted in Figure 4.2) shows that the obtained model accuracy of the proposed model is around 8% better than the Rusentiment proposed by Rogers Rogers et al.

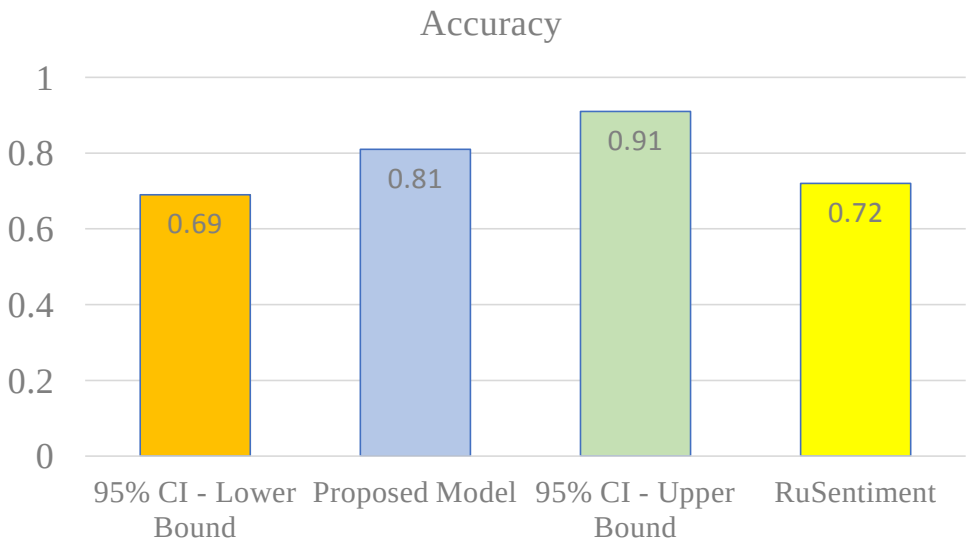



Figure 4.2: Model Accuracy Comparison

(2018). The Rusentiment model is only capable of deciding whether a sentiment is positive or negative regardless of the privacy score level of that sentiment. It also did not consider the background information of users to see if there is any shared structured data related to the sentiment or not. Besides, a case study with the *LIWC* software is analysed to find out the risky sentiments of the proposed model and evaluate it. It has been observed that the privacy risks for 15 sentiments out of 50 were zero. Figure 4.3 shows a comparison of 20 randomly selected sentiments between the proposed model and the *LIWC* software.

As can be seen, in some cases, the privacy score calculated by the software is equal to zero such as for sentiments one and two. The proposed fuzzy model generally provides a higher privacy score compared with the software as it considers structured data in calculation. For better clarification, four different sentiments with four different patterns have been compared. For the first case, the privacy score of the second sentiment from the proposed model is high while the calculated risk from the software is none. (It is not considered risky at all.) The second sentiment has shared the following sentiment: ”Obamacare: Full of Higher Costs and Broken Promises”. Obviously, the shared sentiment may contain sensitive

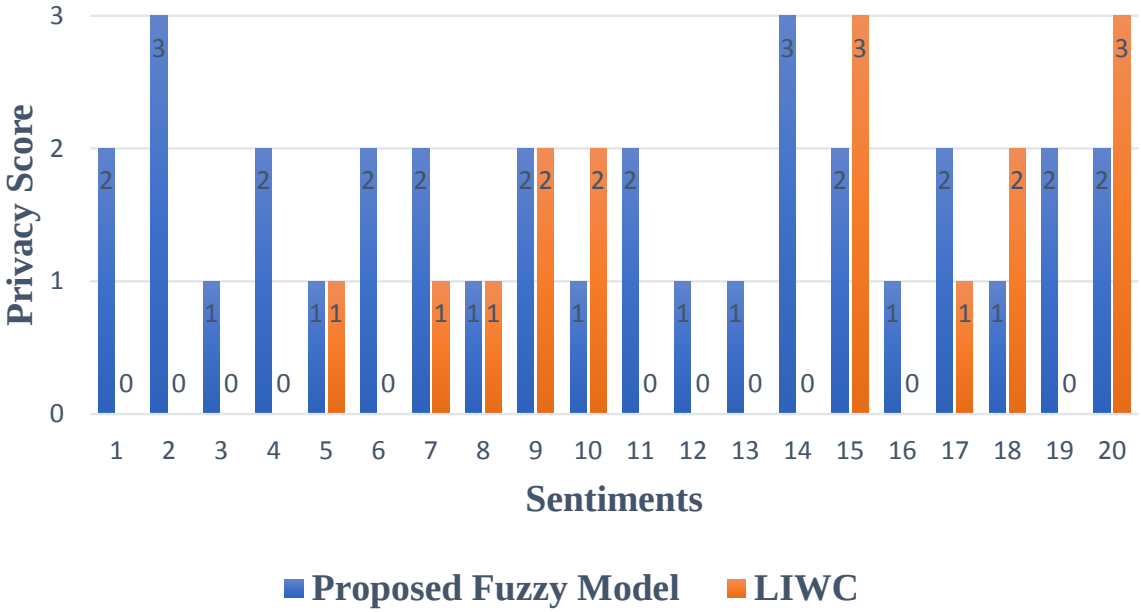



Figure 4.3: Comparison of Proposed Fuzzy Model and LIWC for Privacy Score Calculation
0=No Privacy Risk, 1=Low, 2=Medium, 3=High

information which can bring privacy risks for the user who tweeted the above sentiment. As the user provided his structure data related to the sentiment, then the information amalgamation provides a high privacy score for that sentiment.

For the second case, the proposed model provides the same privacy score as the software does, as depicted for sentiment five. The fifth sentiment is as follow: "Amazon delivery drones show need to update law to promote innovation; protect privacy." As can be seen, the sentiment is expressing the need for improvement in the Amazon company and, as it does not contain any other information, the privacy risk would be low. It should be noted that no structured information exists related to the user who has shared the sentiment.

For the third case, the privacy score of the seventeenth sentiment from the proposed model is medium while the calculated risk from the software is low. The shared sentiment is as follow: "Since 1970 our spending has grown 288% while the median income has grown only 24%– a difference of 264%". Although

the sentiment itself does not come with high sensitivity, but the availability of the background information of the user who shared the sentiment leads to a higher privacy score in the proposed model.

In the last case, the software has provided a higher privacy score for the twentieth sentiment than did the proposed model. The shared sentiment is as follow: "editors agree the President's admin earned their reputation for obfuscation; disregard for failures". As long as there is no background information about the editors, the estimation of the privacy score of the proposed model would not be high, while the software provides a high estimation. By having a look at the sentiment, it can be seen that it does not come with high sensitive information. Hence, medium prediction could be a fair estimation of the privacy risk. As depicted from the figure, in four cases, the software has provided a better estimation for the privacy.

The main reason returns to the availability of background information related to the shared sentiments. As background information of the users plays an important role in fuzzy calculation, lacking this feature may decrease the accuracy of the prediction. Furthermore, the applied dictionaries in the software, compared with the ones obtained from *Wordnet*, are different. Hence, the extracted features from the sentiments may vary which could lead to a dissimilar privacy score. Moreover, as the software is using the pre-defined privacy dictionary for risk scoring, some terms which can influence the privacy might be neglected by their built-in model and hence, the accuracy of the model would collapse. It is worth mentioning that the obtained results from the model and its correctness are validated by the experts in the domain. In a nutshell, it can be seen that the existence of background information from a user who shares a sentiment, as well as information amalgamation from different social networking sites, can provide a more accurate estimation of a privacy score for the shared unstructured data within online social networking sites.

## 4.4 Comparison with Other Techniques

There are different considerations to score the privacy risk of unstructured data for users of online social media sites. While most present research employs machine learning (ML) to analyse the textual data, some scholars applied natural language processing (NLP) techniques to achieve privacy risk calculation. Although NLP can provide a suitable approach to process and analyse the textual data for privacy measurement, there is a need for more effective solutions to identify the sensitive information which can be extracted from unstructured data. Here, the most common approaches for text mining, privacy scoring and anonymisation of unstructured data are discussed.

### 4.4.1 Text Mining

Unstructured data is being shared as various types. The most common type of contextual data is medical records while individuals share different textual data in forums, social networks and blogs. Extracting information from this type of data is one of the significant chores of text mining and has been broadly studied in several research projects such as web mining, information retrieval and natural language processing. The overall goal is to determine the structured information from the information which is semi or fully unstructured. To achieve this, there is a need to focus on two rudimentary principles, namely, relation extraction and named-entity recognition (Jiang, 2012).

The aim of name-entity recognition is to locate references to particular items in texts which are in natural language (Kanya and Ravi, 2012). (Natural language differs from computer codes or artificial language). Support vector machines (Isozaki and Kazawa, 2002), hidden Markov models (Bikel et al., 1997), conditional random fields (CRF) (Settles, 2004) and maximum entropy Markov models (Chieu and Ng, 2002) are the most common machine learning approaches

which are applied for named-entity recognition. Relation extraction is the other fundamental principle in information extraction. The approaches to achieve relation extraction include kernel methods (Zelenko et al., 2003), weakly (semi-) supervised learning and feature-based classification (Aggarwal and Zhai, 2012). Besides the mentioned methods which are all included in supervised information extraction, another method to accomplish the information extraction is unsupervised learning. Contrary to the supervised learning, no priory output would exist, and this technique is only tailored with observations (Janasik et al., 2009). Figure 4.4 shows the taxonomy of current methods for extracting information from text documents.

### 4.4.2 Privacy Scoring and Anonymisation of Unstructured Data

Various forms of texts and unstructured data are shared in a public manner which can bring privacy risks for individuals. While individuals may not be aware of the consequences of sharing text publicly, there should be methods for scoring the privacy of such unstructured data. Different methods have been proposed to score the privacy of textual data. Srivastava and Geethakumari (2013) proposed 'Privacy Armor' to calculate the privacy leaks which would alert the users if they have shared any sensitive unstructured information. They dealt with a response matrix considering the measurement of text messages in a single source of data. They developed a naive quotient model for calculating privacy by assigning binary values for shared and not shared information about the users' profiles, respectively. Their privacy model measures two factors: the sensitivity of the information and the visibility of the information. Their model only considers the structured information within a shared sentiment which cannot suffice to score the privacy of users. Other scholars proposed different methods such as Bag of Words (space vector model), Naive Bayes, Support Vector Machine (SVM) and

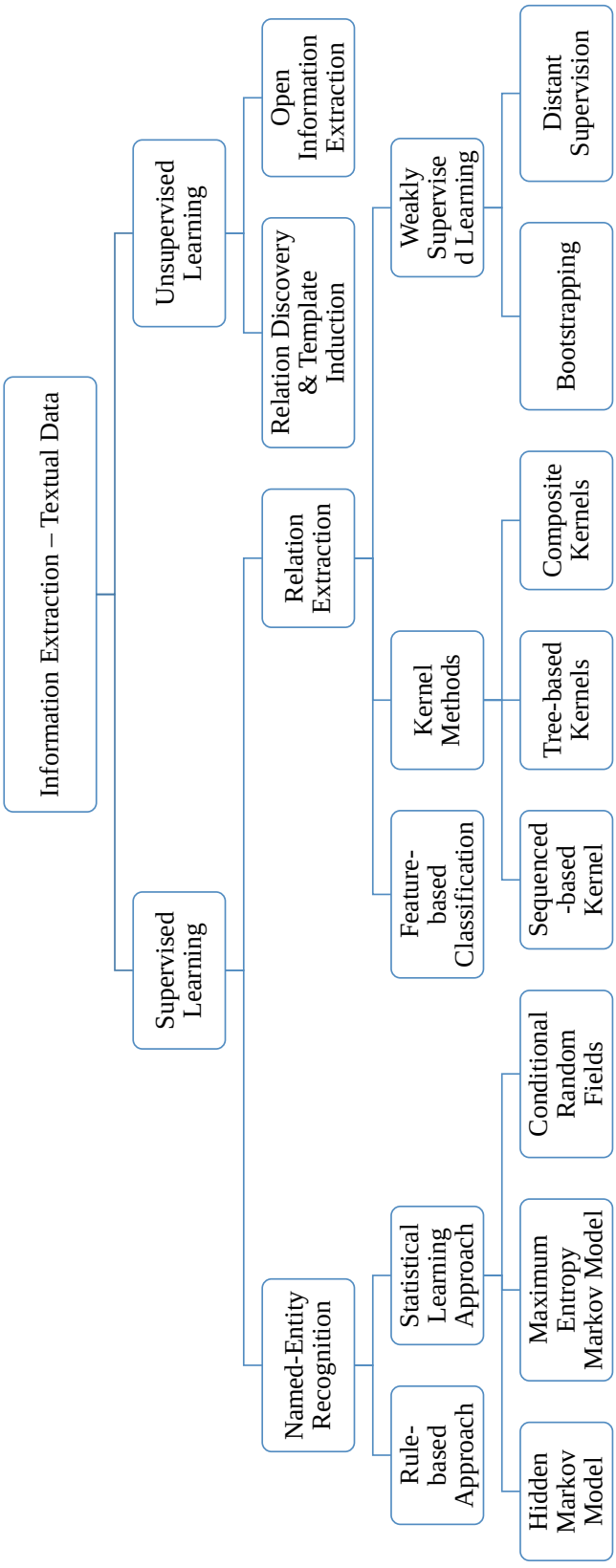



Figure 4.4: A Taxonomy of Information Extraction from Text Data

maximum entropy (Farzindar and Inkpen, 2015; Wilson et al., 2005) to find out the polarity of sentiments, regardless of privacy risks for users. Kumar and Sebastian (2012) proposed a model for determining sentiment's polarity based on the opinion indicator strength but they disregarded the privacy risk. To achieve this, they applied dictionary based and corpus based methods as well as POS-tagging and noise removal to discover the polarity of the sentiments. As the strength of the terms was assigned based on the intuition, it cannot be guaranteed that the provided results for the polarity are accurate. Canfora et al. (2018) proposed an NLP-based model to avert information leak and guarantee the privacy in social media. By considering the grammatical rules of sentiments applying heuristic NLP, they identified sensitive information within a piece of unstructured information shared on social media. They applied an engine which was able to parse the *XML* files where they have defined the heuristics. While their model comes with an acceptable level of accuracy to predict privacy leakage, the privacy of sentiments with similarities may not be accurately classified. Khazaei et al. (2018) discussed the privacy dichotomy for the Twitter profiles considering users' privacy preferences. By analysing the adjacent neighborhoods of users who share sentiments in social networks, they have proved that the privacy preferences in social media are localised. Although scholars proposed ways to understand the polarity of the shared unstructured data or the amount of privacy leakage, none of the above has considered the structured data which is aligned with shared unstructured data. Also, they did not consider the amalgamation of the information from multiple sources of data which can make their models more credible. Hence, there is a need to present a model that can measure the privacy risk of users who share unstructured information on online social media, which provides a new approach to this field when the privacy settings of the social media cannot provide appropriate privacy preservation mechanisms for unstructured data.

## 4.5 Summary

This chapters highlights the difficulties of privacy measurement for unstructured data on online social networks, as well as providing a scoring model. To score the privacy, two measures are considered to score the privacy risk of unstructured data shared by social media users, namely, a machine learning model and a fuzzy based system. In the first stage of calculation, the features from the dataset were extracted in order to comprehend what features can influence on privacy of the users. Later, a rule-based fuzzy system considering the background knowledge of the users is defined to calculate the final privacy score. Regarding the obtained features and results from the two measures, the conclusion is that sentiment privacy is strictly dependent on the number of features which exist in each sentiment and on the existence of the related background information. Finally, the obtained results lead to the conclusion that the proposed model to score the users' unstructured data can suggest a helpful insight for users, enabling them to have a more comprehensive inspection of the information they desire to share in the future. In the next chapter, the way in which a privacy-preserved friending can be undertaken for users who participate in multiple social networking sites is shown.

# Chapter 5

# Privacy-Enhanced Friending Approach for Users on Multiple Online Social Networks

*"I am not a fan of Facebook or Twitter. They both allow too much information to be available and they make privacy a thing of the past." (Kirsty Gallacher)*

In the previous chapters, methods for privacy measurement in multiple online social media for both structured and unstructured data were proposed so that users can make judgments about their level of risk, or allow them to change their use of those platforms. This chapter addresses the challenges of sharing information in a safe manner with unknown individuals. Currently, there are a number of available methods for preserving privacy in order to friending (the act of adding someone as a friend), but they only consider a single source of data and are more focused on users' security rather than privacy. Here, a new privacy-preserving friending method is proposed that helps users to decide what to share with other individuals with reduced risk of being exploited or re-identified.

## 5.1 Introduction

One of the main features of social media sites is friending.[1] Lewis and West (2009) define the friending process as the likelihood of social contact increasing when individuals gather and add other individuals (friends) on a mutual basis to their online social network. Friending helps users to establish links with other users who may be unknown or unmet to the user and allow them to access the content of those users' profiles. Friending on social network sites typically confers a user's particular rights (Thelwall, 2008), besides allowing access to other users' information. Having access to the shared information of these individuals may lead to the misuse of user's information and bring several privacy risks and vulnerabilities to the user. As more information is shared, the probability of privacy risks increases (Houghton and Joinson, 2010; Aghasian et al., 2017).

Recent research investigated a number of friending methods on online social networks. Different scholars (Zhang et al., 2013, 2015b; Preibusch and Beresford, 2009) have investigated privacy preservation mechanisms for friending, which are more focused on secure communication channels for data transmission to achieve data security while they do not concentrate on data privacy of users in the process of friending. Others (Pensa and Di Blasi, 2017; Xu et al., 2017; Veiga and Eickhoff, 2016) have proposed privacy scoring systems without considering mitigation methods to preserve privacy across multiple sources of data. Since online social networks provide an environment to share an individual's information with other users for the purpose of befriending, these current methods cannot be applied to the process of friending with a reduced risk in privacy. Also, current methods do not take into consideration what should be shared in multiple online social network sites to meet users' online privacy needs and what types of information should be preserved. Hence, users need an appropriate system to help

[1] Merriam-Webster defines friending as: to include (someone) in a list of designated friends on a person's social networking site.

them decide what and how information should, or should not, be shared publicly prior to friending, which in turn can preserve their online privacy.

One of the main factors to mitigate the privacy risks and improve the friending process is understanding what types of information are more sensitive and have more impact on privacy for users, and differentiating this information from non-sensitive information. This can provide a better insight into what users can safely share prior to friending on online social networks while still maintaining sufficient privacy. As sharing information on multiple sites increases the probability of a user's information being exploited (Aghasian et al., 2017), the proposed method reduces users' privacy risks for the purpose of friending, considering multiple online social networking sites.

As mentioned in Chapter 3, sensitivity is one of the main factors that influences users' privacy. To measure the sensitivity, a calculation scheme is proposed in order to decide what types of information (that is, attributes) can bring privacy threats to users. This is undertaken by the evaluation of risks to which users may be exposed to, which is accomplished by the Bernstein polynomial function. Additionally, a new anonymisation method based on the sensitivity of users' information on multiple online social networking sites is developed, which assists users to decrease their privacy risks in cyberspace, as well as the privacy of the information they share, in order to achieve friending with others, solicited or unsolicited. Moreover, the proposed framework lets users have more friends in their networks while their privacy is preserved.

The rest of this chapter is organised as follows. Section 5.2 presents the problem definition and methodology of the study. Section 5.3 describes the privacy-preservation framework. Section 5.4 presents the experimental evaluation of the proposed method. Section 5.5 provides a comparison of preserving privacy friending techniques with the proposed model and information sharing privacy on social networking sites. Finally, the proposed model is summarised.

## 5.2 Problem Definition and Methodology

To reduce the risks of the friending process in social networks, there is a need for techniques that consider which sorts of information can lead to privacy risks for users in the friending process and how to provide privacy preservation for them in the case of friending. These techniques also need to be fast in online social media to preserve the privacy of individuals in a timely manner.

Figure 5.1 presents the overall framework for the privacy-enhanced friending technique for online social network users. The framework consists of three main phases: data preparation, sensitivity calculation and the anonymisation process, and finally, the mechanism output. In the data preparation phase, 20 attributes[2] of users are considered and gathered from social networking sites. These attributes are the most common one which were considered in previous studies relating to privacy measurement in online social networking sites. The values for the attributes are obtained from synthetic data, generated using Mockaroo.[3] The software creates realistic-looking data which is close to the real shared information on online social sites. Moreover, the use of testing data makes the results more robust as it provides fewer errors compared with the real data.

The level of sensitivity of attributes is obtained from the user's perspective and indicates how concerned a user is about his/her information being shared publicly. Regarding a user's attributes, there are various classifications by different authors (Årnes et al., 2011; Ho et al., 2009; Richthammer et al., 2014). Årnes et al. (2011) classified users' profile data into three categories consisting of mandatory, extended and personal data. Richthammer et al. (2014) extended the Arnes classification by covering more attributes in social networking sites in order to

[2]The compulsory attributes include name, surname, date of birth, gender and joined date and the rest of attributes gathered from historical studies include college name, company name, school, University, city, state, language, qualification, job position, phone number, email, religious views, political views, interests and postcode.

[3]Mockaroo (https://www.mockaroo.com/) randomly generates test data with requested characteristics.

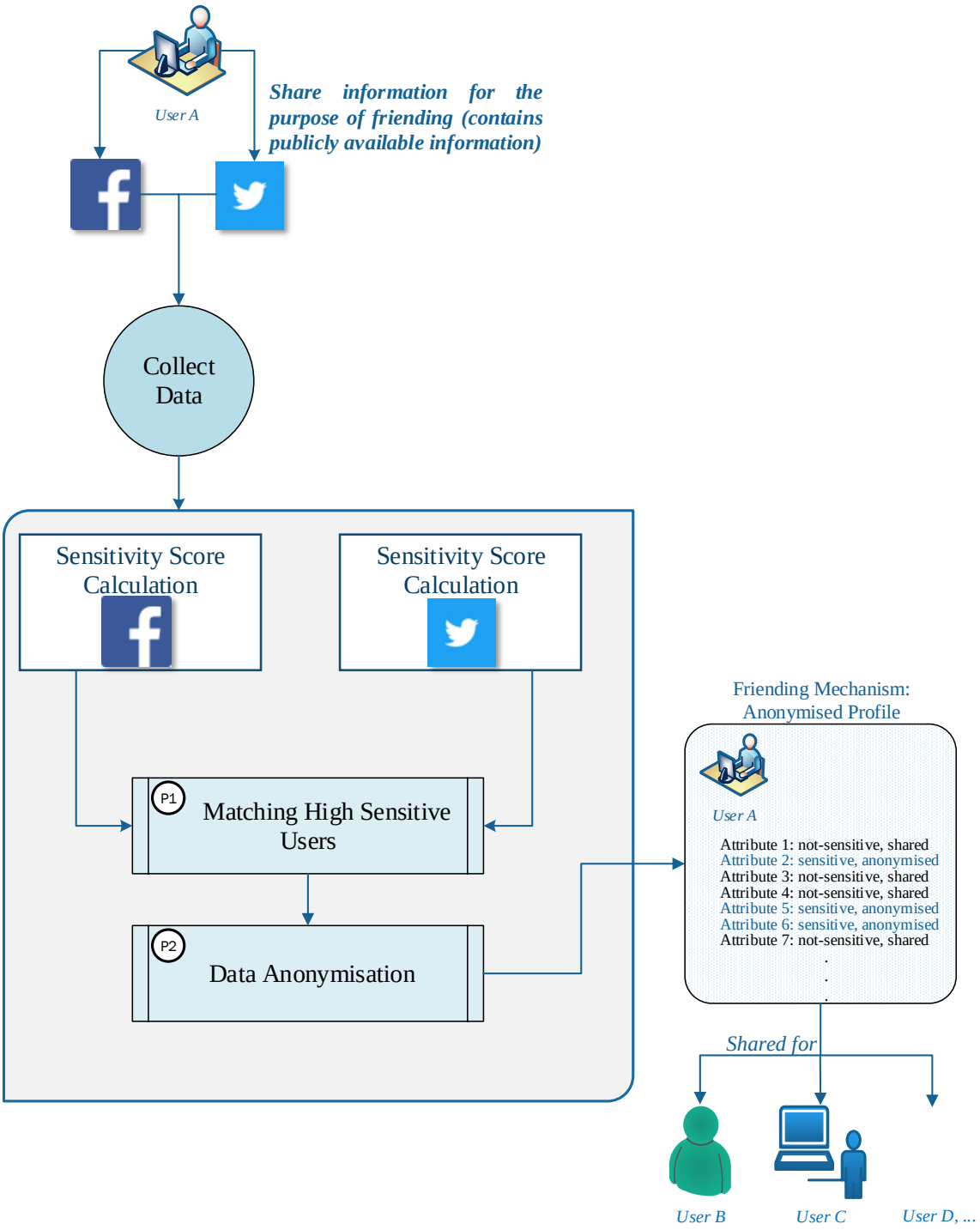



Figure 5.1: Overall Framework of the Privacy-Enhanced Friending Technique

classify the data. Ho et al. (2009) proposed five different groups to classify users' data. Based on the different methods for classifying the users' data in social networking sites, the information of users is categorised into three different levels which consist of personal information of users, compulsory information and sensitive information. Compulsory attributes are the ones which a user should provide before creating an account on social network sites. Sensitive information is the one which subjects to stricter legal requirements for collection, storage use and disclosure such as race, ethnics, political or religious opinions, sexual preferences and memberships in different associations. The rest of the attributes belong to the personal information category.

After classifying the attributes, in the second phase, firstly, the sensitivity score of users' profiles on two online social network sites, Facebook and Twitter is calculated. The sensitivity calculation can show which attributes are more sensitive compared with other attributes. Then, two different processes are identified to reduce the privacy risk of users, labelled *P1* and *P2*. In the *P1* phase, the sensitive information in both social networks and the importance of information for users are recognised. In this phase, users who share sensitive data in both social networks, that is, users who are more susceptible to privacy risk, are found. The aggregation of profiles can help in gaining more data as there are multiple sources of data rather than a single source; while matching the sensitivity results can verify if the profile on online social media belongs to a real identity or a fake profile. After finding users who share sensitive information on both social networking sites, in phase *P2*, the anonymisation method will be applied to the obtained dataset from phase *P1* to help users to preserve their privacy prior to sharing their information for the purpose of friending. In the last phase, based on the sensitivity result, each attribute that may lead to identify an individual directly or is considered as highly sensitive, will be detached. Other attributes that have less sensitivity based on the user's perspective are replaced with fewer semanti-

cally values to decrease the privacy risks for users. In this phase, an anonymised profile of a user is obtained in which the user can securely but publicly share information with other users in order to achieve friending.

## 5.3 Privacy-enhanced Friending Framework

In this section, the proposed methods for calculation of the sensitivity and data anonymisation are presented. Firstly, the sensitivity calculation formulae which have been derived from the Bernstein polynomial model is presented. Next, an anonymisation method which can preserve the privacy of users of social media sites is shown.

### 5.3.1 Sensitivity Score Calculation

As mentioned, sensitivity is defined as the amount of potential privacy risks for users in the case of sharing information on their online presence. Calculating the sensitivity level could help users to have a better understanding of the importance of shared information and assist them to decide what to share with others. Here, five different levels for users' information sensitivity are categorised. The users' information can be (a) extremely sensitive, (b) very sensitive, (c) moderately sensitive, (d) low sensitive and (e) very low sensitive. For measuring the sensitivity of each profile item, it is assumed that each user participates in $k$ different online social networking site. To achieve sensitivity calculation, a new method based on a Bernstein polynomial (Schilling et al., 2012) has been proposed. By applying this model, the sensitivity score of social network profiles can be calculated. (Nava et al. (2012) proved that the Bernstein polynomial is more efficient in a computational manner, is universal and is dominant in computation time compared with traditional polynomials. Hence, this approach has

been applied to measure the sensitivity of information in the study and works well in approximation). The response matrix for sensitivity calculation contains the sensitivity of attributes for each user profile. Users' preferences for the sensitivity of their data are aggregated into this matrix. The value differs based on the nature of attributes, whether the attribute is sensitive, compulsory or personal (how much it is considered sensitive by users). The first step is to compute the sensitivity for each profile item $i$ using the following formula:

$$\theta_i = \frac{N - \left( \sum_{c=1}^{m} R_{i,c} / \sum S_i \right)}{N} \tag{5.1}$$

where $\theta_i$ is the sensitivity score of profile item $i$, $N$ is the number of users, $l$ is equal to the number of rows, $R_{i,c}$ is the summation of each attribute score in the response matrix, $S_i$ is the sum of the sensitivity score of each of the response matrices and $m$ is equal to the scale of the sensitivity ($m$=1,2,3,4,5).

$$R_{i,c} = \sum_{j=1}^{l} \begin{cases} 1 & \text{if } R_{ij} \geq c \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

At the next step, the sensitivity of each profile for each attribute is calculated based on the information sensitivity of each user with the following formula:

$$O_{i,j} = \frac{e^{(\theta_i - S_i)}}{1 + \sum_{j=1}^{m} e^{\sum_{c=1}^{j} (S_i - (Index_i R_{i,c}))}} \tag{5.3}$$

where $j$ indicates the scale of the sensitivity of each attribute for each user which is between $[1, 5]$ and $m$ is equal to ($m$=1,2,3,4,5). $Index_i R_i$ indicates that the individual sensitivity score in the response matrix (from $R_1$,..,$R_m$) and $e$ is the Euler's number. This formula is derived from the polytomous Rasch partial credit score (Masters, 1982). After calculating the response matrix of each value in the

defined scale, each item has a profile sensitivity value in the response matrix $R$. The final sensitivity score for each user profile is derived from a linear combination of a basic linear Bernstein polynomial (Schilling et al., 2012). The basic Bernstein polynomial formula is given by:

$$B_n(x) = \sum_{v=0}^{n} \beta_v b_{v,n}(x) \tag{5.4}$$

where $n$ indicates the degree of polynomial and coefficient $\beta_v$ is called the Bernstein (Bezier) coefficient. Here, this formula is used as a base formula for computing the sensitivity score of each user. The value $\beta_v$ is formulated to the sensitivity of each user, and the value $b_{v,n}(x)$ is formulated to the incremental polynomial of the value of sensitivity for the scale range $[1, 5]$. Hence, Bernstein formula is formulated in order to calculate the sensitivity score value as:

$$F_n = \sum_{i=1}^{n} \theta_i \times \left( O_{i,j}^{j} \times (1 - O_{i,j})^{n-j} \right) \tag{5.5}$$

### 5.3.2 Finding Highly Sensitive Users

After calculating the sensitivity of users' profiles on each social network site, it is necessary to understand which users share sensitive information on both social networking sites to meet the assumption. For doing so, the average sensitivity score of all users who have shared their information on social networks is calculated.

$$Avgscore = \frac{\sum_{i=1}^{n} F_n}{n} \tag{5.6}$$

where *Avgscore* indicates the average score of sensitivity for each online social network sites. Then, the users who share their sensitive information in both sites are found- users who have the sensitivity score higher than the average in both

online social networking sites. In this case, the credibility of the users' information can also be guaranteed.

### 5.3.3 Data Anonymisation

To achieve users' privacy, the sensitive information of users should be preserved. In the friending process, the question of whether or not a user's information can be shared as publicly available is answered. Generalisation and suppression can assure a suitable level of privacy for individuals as they are diversifying values and can help to improve friending results. In phase *P2*, an anonymisation model which can provide online privacy for users in terms of friending is developed. The proposed method contains the following steps. Initially, a table which is comprised of sensitive information on both social network site profiles is considered as an input for the anonymisation process. This table is denoted by a matrix. Then, attributes which have a high or very high sensitivity based on the users' perspective are considered for the calculation.

In the proposed method, based on the sensitivity of attributes, different techniques such as suppression, generalisation (which replaces the value with a less specific semantically consistent value), fuzzy-based rule generalisation and binarisation are applied to provide a consistent anonymised table. For example, in the proposed model, *age* is generalised based on the fuzzy-based rule, while *qualification* is binarised based on the true and false criteria. *University* is followed by a grouping model which links a specific user to a particular university whereas *job* is the outcome of specialisation and generalisation.

### 5.3.4 Time Complexity Comparison

This section provides a comparison of time complexity between the most well-known methods and the proposed method for anonymisation to validate the

Table 5.1: Algorithmic Complexity Comparison

| Algorithm | Order | Privacy model |
|---|---|---|
| Bottom-up (Xu et al., 2006) | $O(n^2 \log n)$ | $k$-anonymity |
| Top-down greedy (Xu et al., 2006) | $O(n \log n)$ | $k$-anonymity |
| Mondrian (LeFevre et al., 2006) | $O(n)$ | $k$-anonymity |
| Clustering-based (Lin and Wei, 2008) | $O\left(\frac{n^2}{k}\right)$ | $k$-anonymity |
| The proposed method | $O(n^2 \log n)$ | Sensitivity calculation and $k$-anonymity |

model. As the complexity ($O$) is hardware-independent, this comparison is free from any implementation bias. Table 5.1 compares the computation cost of the proposed method with the most well-known anonymisation methods. Among all anonymisation methods, Le Fevre's Mondrian algorithm (LeFevre et al., 2006) is the fastest local method among these approaches without considering sensitivity of information and multiple sources of information. But the proposed method applies to both the sensitivity calculation and the anonymisation technique. In the model, looking for users to see who obtains high sensitive profiles on both social networks or not is $O(n^2)$. The anonymisation process which contains fuzzy rules, generalisation and suppression has $O(n \log n)$ complexity. Hence, the aggregated complexity for the proposed method has $O(n^2 \log n)$ complexity. While the bottom-up method has the same order compared with ~~theour~~ proposed method, it only focuses on data anonymisation in a single source without calculating the sensitivity of the attributes.

## 5.4 Experimental Evaluation

In this section, an assessment of the proposed privacy-enhanced friending framework is presented. Firstly, the results of the sensitivity score are presented. Then, the results obtained from the proposed anonymisation method are shown.

### 5.4.1 Sensitivity Score

In this section, the sensitivity score of 100 users on two different online social networks, Facebook and Twitter, are provided. The selected number of individuals covers a variety of values from users' social profiles that is desirable to confirm the usefulness of the proposed privacy-enhanced friending model. Based on the categorisation and the sensitivity of attributes from the literature, a random number for the attributes between 1 and 5, considering the discrete uniform distribution is assigned. In this case, a random number between 3 and 5 is assigned to the attributes which are more sensitive and, for the rest of the attributes which are compulsory or personal, a value between 1 and 5 is assigned.

Figures 5.2a, 5.2b and 5.3 show the comparison of sensitivity score of users in three different categories (very high to very low) for these synthetic users. Table 5.2 shows the bounds of sensitivity for each category and the final sensitivity score to determine the level of sensitivity. As mentioned before, three different categories have been considered for the attributes. To define the lower bound and upper bound of each category, the mean function for the attributes related to each category in each social networks is applied. Next, the difference of the obtained result is calculated and divided it into five equal boundaries to determine the five different scales for the categories. At the next stage, the overall sensitivity for each category was calculated and compared with the combined dataset which contains the information of users who have a high sensitivity score in both social networking sites. As it can be seen from Figures 5.2–5.3, analyses of synthetic

Table 5.2: Sensitivity of Information Bounds

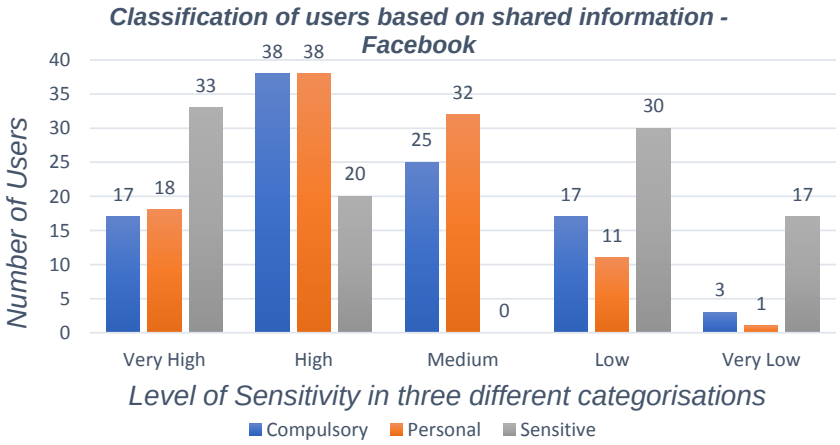
| Sensitivity | Facebook | | | | Twitter | | | |
|---|---|---|---|---|---|---|---|---|
| | C | P | S | F | C | P | S | F |
| VL - Upper bound | 0.126 | 0.195 | 0.024 | 0.29 | 0.18 | 0.16 | 0.03 | 0.19 |
| VL - Lower bound | 0.0104 | 0.163 | 0.02 | 0.235 | 0.15 | 0.128 | 0.022 | 0.152 |
| L - Upper bound | 0.103 | 0.162 | 0.019 | 0.234 | 0.14 | 0.127 | 0.021 | 0.151 |
| L - Lower bound | 0.083 | 0.13 | 0.015 | 0.18 | 0.109 | 0.1 | 0.017 | 0.115 |
| M - Upper bound | 0.082 | 0.129 | 0.014 | 0.179 | 0.108 | 0.099 | 0.016 | 0.114 |
| M - Lower bound | 0.07 | 0.098 | 0.02 | 0.125 | 0.077 | 0.073 | 0.0.012 | 0.078 |
| H - Upper bound | 0.06 | 0.097 | 0.01 | 0.124 | 0.076 | 0.072 | 0.011 | 0.077 |
| H - Lower bound | 0.039 | 0.065 | 0.006 | 0.07 | 0.044 | 0.045 | 0.007 | 0.042 |
| VH - Upper bound | 0.038 | 0.064 | 0.005 | 0.069 | 0.043 | 0.044 | 0.006 | 0.041 |
| VH - Lower bound | 0.017 | 0.032 | 0.001 | 0.011 | 0.011 | 0.017 | 0.001 | 0.004 |

Risk categories are Very low (VL), (L)ow, (M)edium, (H)igh and Very high (VH). Information categories are (C)ompulsory, (P)ersonal and (S)ensitive; F is the final sensitivity score.

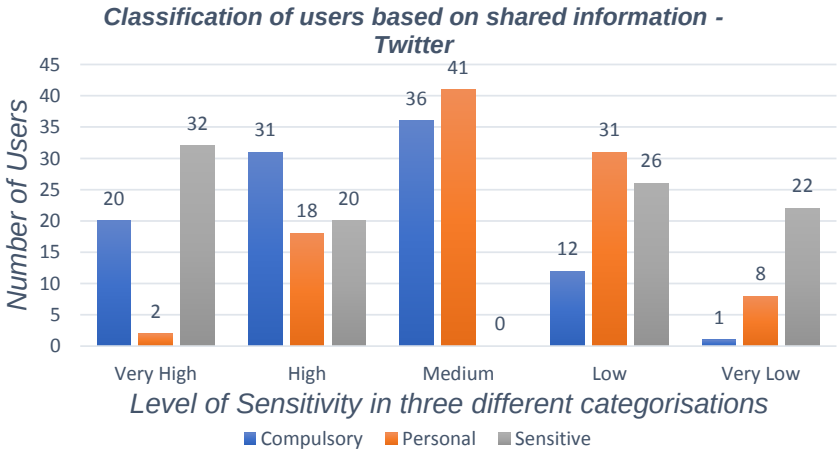
data show that nearly 53% of users do not desire to share their information on online social networks publicly, while the rest of the users have a trade-off between their shared information and their privacy on social media. It can be seen that nearly 22% of users do share their information or they may have accidentally shared information on such social sites which are publicly available.

***Note:*** In this study, it is only measured how sensitive are the values of attributes for the users when they share them in online social networks. Measuring the level of users' knowledge about privacy is out of the scope of this study.

The sensitivity score of synthetic data of users on social networking sites indicates that the number of users who share sensitive information on social networking sites publicly is almost equal to the final result. As mentioned, the provided test data is robust, hence, it can be concluded that the sensitive information plays a critical role in determining whether or not a user is really in a privacy risk. Meanwhile, other categories and attributes of users which are shared cannot be




(a) Percentage of users who have shared information in three different categories of users' data type in online social networks - Facebook case




(b) Percentage of users who have shared information in three different categories of users' data type in online social networks - Twitter case

Figure 5.2: Users' Information Sharing Sensitivity

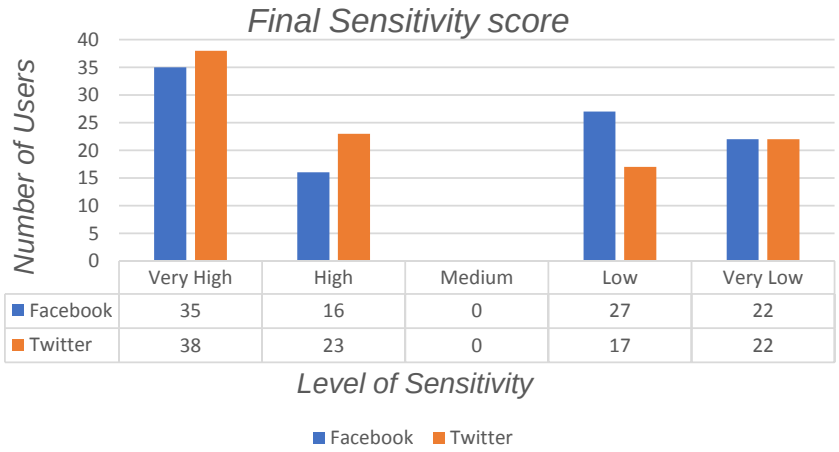



Figure 5.3: Final sensitivity score, for synthetic user data

neglected.

### 5.4.2 Anonymisation Output

In the *P2* phase, the data from matched users obtained from the Bernstein polynomial model are anonymised. Tables 5.3, 5.4 and 5.5 show examples of anonymisation output of the proposed model from a set of raw (synthesised) data. By comparing the results of two tables, users can understand what they can share publicly with other users, what type of information should be shared with fewer semantically values and what type of information should not be accessible to others because of its importance. Hence, the online privacy of users can be met while they have an active presence on social network sites. As a case in point, the automated system can help users to achieve friending with other individuals on online social networks with the risk of privacy mitigated.

## 5.5 Comparison with Other Techniques

While users share information on social networks with friends and even with potential friends in a safe manner, there is a need to present a model to provide

Table 5.3: Raw Data vs. Anonymised Data Publishing: Raw Data Case

| ID | Name | University | Age | Qualification | Job | Postcode |
|---|---|---|---|---|---|---|
| 4 | John | UTAS | 21 | High-school | Secretary | 04526 |
| 5 | Alex | UTAS | 29 | Bsc | Sales | 04572 |
| 6 | Emma | UTAS | 28 | High-School | Marketing | 04637 |
| 7 | Alynn | UTAS | 25 | Msc | Nurse | 04578 |
| 8 | Ho | UTM | 24 | Bsc | Media Planner | 04272 |

Table 5.4: Raw Data vs. Anonymised Data Publishing: Anonymised Data Case

| ID | Name | University | Age | Qualification | Job | Postcode |
|---|---|---|---|---|---|---|
| 4 | * | UTAS | <30 | Non-degree | Secretary | 045-* |
| 5 | * | UTAS | <30 | Degree | Sales | 045-* |
| 6 | * | UTAS | <30 | Non-degree | Marketing | 046-* |
| 7 | * | UTAS | <30 | Degree | Nurse | 045-* |
| 8 | * | UTM | <30 | Degree | Media Planner | 042-* |

Table 5.5: Applied Data Disclosure Method - Sample

| Attribute | Amount of disclosure | Applied Method |
|---|---|---|
| Age | less semantically values | Fuzzy-based rule generalisation |
| Qualification | less semantically values | Binarisation |
| Postcode | less semantically values | Generalisation |

users with their information secrecy and privacy. Information disclosure may lead to privacy risks for the user's friends as well (Alsarkal et al., 2018). Sharing personally identifiable information (PII)[4] can result in several privacy issues for the users such as fraud, stalking, identity theft or perhabs even harassment.

To preserve privacy in terms of friending, several schemes have been proposed which are mostly cryptographic-based and applied for preserving information security but neglecting the privacy of users' contexts on social networks sites. Zhang et al. (2013, 2015b) proposed a secure mechanism for profile matching for the participants who want to be friend. The concentration of their mechanism is on communication security between the friend request sender and po-

[4]Typically, an identifying variable is one that defines an attribute of an individual that is visible and evident, which is recorded (such as social security number, employee ID, patient ID and so on), and which other people can identify.

tential users with whom the user wants to connect. To achieve this, they provide a secure communication channel applying a cryptographic hash of the attributes and the normalisation of the profile. Preibusch and Beresford (2009) investigated the friendship's nature and applied hash identifiers to create a hidden friendship between users by excerpting the friend of friends files between the root and the leaf of the graph they have from the network. Baden et al. (2009) proposed the *Persona* system. By applying the attribute-based encryption (ABE), this system allows individuals to have personalised privacy by applying attribute-based encryption. Guha et al. (2008) proposed a novel approach for preserving privacy based on the proof of concept method which can preserve data privacy, neglecting the fact that the privacy of users is not assured in this system.

Despite various research studies on privacy calculation or privacy preservation, an effective privacy preservation technique for multiple sources of social data considering measurement of highly sensitive information, has not been considered and developed. It is essential to measure privacy when it is distributed across multiple online social networking sites, as more sensitive information of an individual is shared in this way, compared with a single source of data. This can help to perceive which sensitive information should be considered as private (does not go online publicly) when individuals want to have an active online presence, find friends and safely share their information with those whom they want to befriend.

## 5.6 Summary

Providing only a single method of anonymisation for users' data and considering a single social network site is not sufficient to preserve individuals' sensitive information when they want to connect with other individuals. In addition, it will not help users to have a clear understanding of what they can really share with

others individuals when they want to be friends with them. In this chapter, a friending model which applies a mixture of Bernstein theorem and *k*-anonymity anonymisation techniques in multiple online social network sites to preserve the privacy of users is proposed. The proposed friending model can help users to see the level of sensitivity of their information and then provide an anonymised profile of social networks which can provide a better idea about their information sharing behaviour with the privacy-enhanced friending technique prior to friending other users. The next chapter, will summarise the work undertaken in this thesis and its objectives.

# Chapter 6
# Conclusions and Future Directions

*"Security is, I would say, our top priority because for all the exciting things you will be able to do with computers - organizing your lives, staying in touch with people, being creative - if we don't solve these security problems, then people will hold back." (Bill Gates)*

This chapter summarises the objectives and work undertaken in this thesis. The main findings and lessons learned from the study are discussed along with their significance. It concludes with a discussion of future work, the need for which emerged during this research.

## 6.1 Conclusions

As the usage of social media grows daily, privacy-related concerns are becoming more and more important. Online social media users usually have several social network profiles for different purposes. In each network, individuals share sensitive and personal information for different purposes such as friending and communication. Sharing information can lead to privacy risks for an individual who participates in such networks. These risks can be considered from two different viewpoints: calculating a user's privacy risk of accidental information revelation, and methods to safeguard users' privacy when sharing great amount of information.

In the beginning of this thesis, a taxonomy was developed to classify the com-

mon privacy scoring and preservation methods which are applied in online social media networks. The taxonomy provides the basis for comparing different privacy approaches for individuals who participate in social media sites. This comprehensive classification not only builds up the understanding of recent improvements in social media privacy, but it also provides an insight into research gaps that need to be addressed. Based on the literature study, the lack of appropriate mechanisms for privacy scoring and friending for the users who participate in multiple online social networking sites was indicated. From the privacy scoring viewpoint, there is a need to understand what can influence users' privacy on social media sites. From the privacy-preservation viewpoint, the challenges of protecting the privacy of online social media users is to offer techniques which can support individuals to maintain an acceptable privacy level in their online presence.

In Chapter 3, this thesis investigated the level of privacy risk caused by shared *structured* information in multiple online social media sites. It considered three aspects to compute the overall privacy disclosure score of a user who participates in multiple social networks. Firstly, a fuzzy-based system is presented to measure information visibility as a factor that has a direct influence on a user's privacy disclosure score. By comparing the obtained privacy scores of individuals, it was confirmed that users' privacy disclosure scores directly rely on the amount of revealed information of a user, such as interests, job details, and marital status. Secondly, consideration of multiple sources of data, as well as a polytomous approach (different states for data visibility), rather than a dichotomous approach (only public or private visibility state for data), provides a more accurate estimation of the privacy score of users' information. Lastly, the results of the study led to the conclusion that the proposed framework could increase the awareness of users and provide a better perception regarding their online privacy on social media sites.

In Chapter 4, an automated system is proposed to measure the privacy risks caused by shared *unstructured* information within online social networking sites. To achieve this, two measures to score the privacy risk for users were considered, namely, a machine learning model and a fuzzy based system . In the first phase, the features of the textual data are extracted to understand privacy related ones. Next, a set of fuzzy rules was defined to consider the background knowledge of users in order to measure the final privacy score. Considering a combination of two measures increases the accuracy of the privacy estimation more than previously proposed methods. Regarding the obtained results of both phases, it was concluded that the sentiment's privacy is strictly dependent on the number of the features which exists in each sentiment. Moreover, the obtained results lead to the conclusion that the proposed model can provide insight for individuals to develop a more comprehensive review of the information they desire to disclose in the future.

In Chapter 5, a framework to enable users to achieve friending with the reduced risk was developed. Although anonymisation techniques can preserve the privacy of users, but they do not consider the privacy from a user's perspective who shares information on various social media sites. Furthermore, anonymisation alone will not support individuals to have a clear understanding of what information can be shared with other users. Hence, a friending framework using a mixture of different statistical (Bernstein theorem) and anonymisation techniques ($k$-anonymity) has developed which considers multiple online social network profiles of a user. The proposed model can be applied to understand the sensitivity level of users' information and provide an anonymised profile. Hence, users can have a better understanding of their information sharing behaviour using the proposed privacy-enhanced friending model before sending any requests to other social media users. Moreover, the time complexity comparison of the existing methods with the proposed friending mechanism indicates that the pro-

posed method has less computational cost.

In summary, the proposed models in this thesis have taken into account both structured and unstructured data, as well as multiple sources of social media in order to provide a better estimation of the privacy risk for social media users and for privacy preservation.

## 6.2 Future Directions

In this thesis, the problem of information sharing with other individuals in multiple online social media and methods for mitigating the privacy risks of users have been discussed. However, there are still open problems that could be considered for further research in the future.

### 6.2.1 Privacy Score Generalisation for Unstructured Data

Although scoring the risk level of users in social media can help them to understand the level of their privacy, further generalisation of the privacy scoring model considering users' viewpoints about the sensitivity of the terms within a sentence should be explored. Furthermore, there are no criteria for weighting the terms in the privacy calculation of the unstructured data, which is, hence, an open problem for further investigation.

### 6.2.2 Privacy Preservation Modelling

In terms of privacy preservation, new models for unstructured data should be proposed as they are very popular among social media users. As the textual information of users is being mined by third parties, there should be ways to protect users from privacy breaches such as re-identification. Another consideration is resource consumption. As providing privacy requires substantial computation,

factors that impact on computation time should be studied to identify whether or not new mechanisms are required that are less computationally intensive.

### 6.2.3 Attack Modelling

Modelling attack is a significant task that should be taken into account to enable data protection and preservation. While different types of attacks occur on anonymised data, modelling attacks on datasets with more complicated features should be a priority so that vulnerabilities can be uncovered before they are exploited. There is also a need to propose novel methods to assist organisations in the preservation of the privacy of users while these organisations are storing, analysing and mining individuals' data.

### 6.2.4 Privacy Personalisation

As the approaches to the sensitivity and privacy of information are different among users, it is essential to provide personalised privacy for users within online social media sites. Hence, it is interesting to focus on personalisation in privacy preservation in order to increase the data utility of the shared information of individuals. This can help users to benefit from a model that matches their privacy perspectives. Future studies could consider the measurement of users' knowledge about privacy and the selection of privacy settings on social networking sites, and the trade-off between privacy and online presence, as well as improving the personalisation of individuals.

### 6.2.5 Real-Time Privacy Alert System

Existing work on scoring and preserving the privacy of users in social media do not offer a real-time alert system or risk level calculation when a user provides

extra information on his/her profile. Hence, it would be interesting to build a classification system that could help analyse the information of the users and gain more detailed insights about the relationships between shared information in real time. Using other methods of machine learning for solving this problem would also be worth examining in future research.

### 6.2.6 Privacy Functionality Score

One of the approaches that can taken into account it to measure the benefits a user can obtain by participating in online social media sites. Hence, it would be interesting to create functionality scoring systems to help users to make the best use of social media sites while their online privacy is being preserved in such sites.

# References

Abuelsaad, T. E. and Hoyos, C. (2011). Data perturbation and anonymization using one way hash, patents.

Aggarwal, C. C. and Zhai, C. (2012). *Mining Text Data*. Springer Science & Business Media, ISBN:978-1-4614-3222-7, pp.1–521.

Aghasian, E., Garg, S., Gao, L., Yu, S., and Montgomery, J. (2017). Scoring users' privacy disclosure across multiple online social networks. *IEEE access*, 5:13118–13130.

Ahmadizadeh, E., Aghasian, E., Taheri, H. P., and Nejad, R. F. (2015). An automated model to detect fake profiles and botnets in online social networks using steganography technique. *IOSR Journal of Computer Engineering*, 17(1):65–71.

Aldhafferi, N., Watson, C., and Sajeev, A. (2013). Personal information privacy settings of online social networks and their suitability for mobile internet devices. *International Journal of Security, Privacy and Trust Management (IJSPTM)*, 2(2):1–17.

Alsarkal, Y., Zhang, N., and Xu, H. (2018). Your privacy is your friend's privacy: Examining interdependent information disclosure on online social networks. In *Proceedings of the 51st Hawaii International Conference on System Sciences*, pages 892–901.

an, M. E. and Mattord, H. J. (2011). *Principles of information security, Book*. Cengage Learning, isbn:1111138214, pp. 1–609.

Anderson, J. (2013). *Privacy engineering for social networks*. PhD thesis, University of Cambridge.

Årnes, A., Skorstad, J., and Michelsen, L. (2011). Social network services and privacy. *Datatilsynet, Oslo*, page 40.

BACKSTROM, L., DWORK, C., and KLEINBERG, J. (2007). Wherefore art thou r3579x? anonymized social networks, hidden patterns, and structural steganography. *Communications of the ACM*, 54(12):133–141.

Baden, R., Bender, A., Spring, N., Bhattacharjee, B., and Starin, D. (2009). Persona: an online social network with user-defined privacy. In *ACM SIGCOMM Computer Communication Review*, volume 39, pages 135–146.

Badsha, S., Yi, X., and Khalil, I. (2016). A practical privacy-preserving recommender system. *Data Science and Engineering*, 1(3):161–177.

Becker, J. L. (2009). Measuring privacy risk in online social networks. *IEEE Security*, 1:1–8.

Bell, C. (2014). How over sharing on social media can cost you.

Beye, M., Jeckmans, A. J. P., Erkin, Z., Hartel, P., Lagendijk, R. L., and Tang, Q. (2012). *Privacy in online social networks*, pages 87–113. Springer.

Bikel, D. M., Miller, S., Schwartz, R., and Weischedel, R. (1997). Nymble: a high-performance learning name-finder. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*, pages 194–201. Association for Computational Linguistics.

Billsus, D. and Pazzani, M. J. (2000). User modeling for adaptive news access. *User modeling and user-adapted interaction*, 10(2-3):147–180.

Bonneau, J. (2009). Attack of the zombie photos. *Light Blue Touchpaper http://www. lightbluetouchpaper. org/2009/05/20/attackof-the-zombie-photos*.

Bonti, A., Li, M., Gao, L., and Shi, W. (2012). Effects of social characters in viral propagation seeding strategies in online social networks. In *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 632–639. IEEE.

Boutet, A., Frey, D., Guerraoui, R., Jégou, A., and Kermarrec, A.-M. (2016). Privacy-preserving distributed collaborative filtering. *Computing*, 98(8):827–846.

Boyd, D. and Ellison, N. (2009). social network sites: definition, history, scholarship: Department of telecommunication. *Information Studies, and Media, Michigan State University*, 13(1):210–230.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370.

Bünnig, C. and Cap, C. H. (2009). Ad hoc privacy management in ubiquitous computing environments. In *Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services, 2009. CENTRIC'09. Second International Conference on*, pages 85–90. IEEE.

Canfora, G., Di Sorbo, A., Emanuele, E., Forootani, S., and Visaggio, C. A. (2018). A nlp-based solution to prevent from privacy leaks in social network posts. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*, pages 36–44. ACM.

Casino, F., Domingo-Ferrer, J., Patsakis, C., Puig, D., and Solanas, A. (2015). A k-anonymous approach to privacy preserving collaborative filtering. *Journal of Computer and System Sciences*, 81(6):1000–1011.

Cherdantseva, Y. and Hilton, J. (2013). A reference model of information assurance & security. In *Availability, reliability and security (ares), 2013 eighth international conference on*, pages 546–555. IEEE.

Cheung, C. M., Chiu, P.-Y., and Lee, M. K. (2011). Online social networks: Why do students use facebook? *Computers in Human Behavior*, 27(4):1337–1343.

Chew, M., Balfanz, D., and Laurie, B. (2008). (under) mining privacy in social networks. Technical report, Google Inc.

Chieu, H. L. and Ng, H. T. (2002). Named entity recognition: a maximum entropy approach using global information. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1*, pages 1–7. Association for Computational Linguistics.

Counsel, C. P. (2014). Privacy and data protection act. Technical report, Parliament of Victoria.

Domingo-Ferrer, J. (2010). Rational privacy disclosure in social networks. In *International Conference on Modeling Decisions for Artificial Intelligence*, pages 255–265. Springer.

Domingo-Ferrer, J., Viejo, A., Sebé, F., and González-Nicolás, Ú. (2008). Privacy homomorphisms for social networks with private relationships. *Computer Networks*, 52(15):3007–3016.

Dwork, C. (2008). An ad omnia approach to defining and achieving private data analysis. In *Privacy, Security, and Trust in KDD*, pages 1–13. Springer.

Dwork, C. and Rothblum, G. N. (2016). Concentrated differential privacy. *arXiv preprint arXiv:1603.01887*.

Edgar, D. (2000). Data sanitization techniques. Technical report, A Net.

Fang, L. and LeFevre, K. (2010). Privacy wizards for social networking sites. In *Proceedings of the 19th international conference on World wide web*, pages 351–360. ACM.

Farzindar, A. and Inkpen, D. (2015). Natural language processing for social media. *Synthesis Lectures on Human Language Technologies*, 8(2):1–166.

Fiesler, C., Dye, M., Feuston, J. L., Hiruncharoenvate, C., Hutto, C. J., Morrison, S., Khanipour Roshan, P., Pavalanathan, U., Bruckman, A. S., De Choudhury, M., et al. (2017). What (or who) is public? privacy settings and social media content sharing. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 567–580. ACM.

Fulgoni, G. M. (2018). How limited data access constrains marketing-mix analytical efforts: Why data barriers are preventing marketers from optimizing marketing spend. *Journal of Advertising Research*, 58(4):390–393.

Fung, B., Wang, K., Chen, R., and Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (CSUR)*, 42(4):14.

Fung, B., Wang, K., and Yu, P. S. (2007). Anonymizing classification data for privacy preservation. *Knowledge and Data Engineering, IEEE Transactions on*, 19(5):711–725.

Fung, B. C. M., Wang, K., and Yu, P. S. (2005). Top-down specialization for information and privacy preservation. In *21st International Conference on Data Engineering (ICDE'05)*, pages 205–216. IEEE.

Ghanei, S. and Faez, K. (2016). Localizing scene texts by fuzzy inference systems and low rank matrix recovery model. *Computer Vision and Image Understanding*, 142:94–110.

Grabisch, M., Nguyen, H. T., and Walker, E. A. (2013). *Fundamentals of uncertainty calculi with applications to fuzzy inference*, volume 30. Springer Science & Business Media.

Gross, R. and Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the Electronic Society*, pages 71–80.

Guha, S., Tang, K., and Francis, P. (2008). Noyb: Privacy in online social networks. In *Proceedings of the First Workshop on Online Social Networks*, pages 49–54. ACM.

Guo, L., Zhang, C., Fang, Y., and Lin, P. (2015). A privacy-preserving attribute-based reputation system in online social networks. *Journal of Computer Science and Technology*, 30(3):578–597.

Ho, A., Maiga, A., and Aïmeur, E. (2009). Privacy protection issues in social networking sites. In *Computer Systems and Applications, 2009. AICCSA 2009. IEEE/ACS International Conference on*, pages 271–278. IEEE.

Hofmann, T. and Hartmann, D. (2005). Collaborative filtering with privacy via factor analysis. In *Proceedings of the 2005 ACM Symposium on Applied Computing*, pages 791–795. ACM.

Houghton, D. J. and Joinson, A. N. (2010). Privacy, social network sites, and social relations. *Journal of Technology in Human Services*, 28(1-2):74–94.

Isozaki, H. and Kazawa, H. (2002). Efficient support vector classifiers for named entity recognition. In *Proceedings of the 19th international conference on Compu-*

*tational linguistics-Volume 1*, pages 1–7. Association for Computational Linguistics.

Janasik, N., Honkela, T., and Bruun, H. (2009). Text mining in qualitative research: Application of an unsupervised learning method. *Organizational Research Methods*, 12(3):436–460.

Jeckmans, A. J., Beye, M., Erkin, Z., Hartel, P., Lagendijk, R. L., and Tang, Q. (2013). Privacy in recommender systems. In *Social media retrieval*, pages 263–281. Springer.

Jiang, J. (2012). Information extraction from text. In *Mining Text data*, pages 11–41. Springer.

Jorgensen, Z. and Yu, T. (2014). A privacy-preserving framework for personalized, social recommendations. In *EDBT*, pages 571–582.

Kafalı, Ö., Günay, A., and Yolum, P. (2014). Detecting and predicting privacy violations in online social networks. *Distributed and Parallel Databases*, 32(1):161–190.

Kang, J. (1998). Information privacy in cyberspace transactions. *Stanford Law Review*, 50:1193–1294.

Kanya, N. and Ravi, T. (2012). Modelings and techniques in named entity recognition: an information extraction task. *3rd International on Sustainable Energy and Intelligent Systems*, pages 104–108.

Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of social media. *Business horizons*, 53(1):59–68.

Kenny, G. and Connolly, R. (2016). Drivers of health information privacy concern: A comparison study. In *HEALTHCARE INFORMATICS AND INFORMATION TECHNOLOGY*, pages 1–10.

Khazaei, T., Xiao, L., Mercer, R. E., and Khan, A. (2018). Understanding privacy dichotomy in twitter. In *Proceedings of the 29th on Hypertext and Social Media*, pages 156–164. ACM.

Kowalski, R. M. (2000). I was only kidding victims and perpetrators perceptions of teasing. *Personality and Social Psychology Bulletin*, 26(2):231–241.

Kumar, A. and Sebastian, T. M. (2012). Sentiment analysis on twitter. *International Journal of Computer Science Issues (IJCSI)*, 9(4):372.

Leenes, R. (2009). Context is everything sociality and privacy in online social network sites. In *IFIP PrimeLife International Summer School on Privacy and Identity Management for Life*, pages 48–65. Springer.

LeFevre, K., DeWitt, D., and Ramakrishnan, R. (2006). Mondrian multidimensional k-anonymity. In *Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on*, pages 25–25. IEEE.

LeFevre, K., DeWitt, D. J., and Ramakrishnan, R. (2005). Incognito: Efficient full-domain k-anonymity. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 49–60. ACM.

Lenhart, A. (2008). Teens, online stranger contact & cyberbullying: What the research is telling us. Technical report, Pew Internet & American Life Project.

Lewis, J. and West, A. (2009). 'friending': London-based undergraduates' experience of facebook. *New Media & Society*, 11(7):1209–1229.

Li, D., Chen, C., Lv, Q., Shang, L., Zhao, Y., Lu, T., and Gu, N. (2016). An algorithm for efficient privacy-preserving item-based collaborative filtering. *Future Generation Computer Systems*, 55:311–320.

Li, M., Yu, S., Zheng, Y., Ren, K., and Lou, W. (2013). Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. *Parallel and Distributed Systems, IEEE Transactions on*, 24(1):131–143.

Li, N., Li, T., and Venkatasubramanian, S. (2007). t-closeness: Privacy beyond k-anonymity and l-diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pages 106–115. IEEE.

Li, N., Li, T., and Venkatasubramanian, S. (2010). Closeness: A new privacy measure for data publishing. *IEEE Transactions on Knowledge and Data Engineering*, 22(7):943–956.

Li, T., Li, N., Zhang, J., and Molloy, I. (2012). Slicing: A new approach for privacy preserving data publishing. *IEEE transactions on knowledge and data engineering*, 24(3):561–574.

Lin, J.-L. and Wei, M.-C. (2008). An efficient clustering method for k-anonymization. In *Proceedings of the 2008 International Workshop on Privacy and anonymity in information society*, pages 46–50. ACM.

Lipschultz, J. H. (2014). Social media communication: Concepts, practices, data, law and ethics. Technical report, Amazon.

Liu, K. and Terzi, E. (2010a). A framework for computing the privacy scores of users in online social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(1):6.

Liu, K. and Terzi, E. (2010b). A framework for computing the privacy scores of users in online social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(1):6.

Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007).

l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3.

Machanavajjhala, A., Korolova, A., and Sarma, A. D. (2011). Personalized social recommendations: accurate or private. *Proceedings of the VLDB Endowment*, 4(7):440–450.

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., and Beaton, M. (2013). Teens, social media, and privacy. *Pew Research Center*, 21.

Madden, M. and Smith, A. (2010). Reputation management and social media. Technical report.

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2):149–174.

Maximilien, E. M., Grandison, T., Liu, K., Sun, T., Richardson, D., and Guo, S. (2009). Enabling privacy as a fundamental construct for social networks. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, pages 1015–1020. IEEE.

Mazzia, A., LeFevre, K., and Adar, E. (2012). The pviz comprehension tool for social network privacy settings. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 13. ACM.

Mendel, J. M. and Wu, H. (2006). Type-2 fuzzistics for symmetric interval type-2 fuzzy sets: Part 1, forward problems. *IEEE Transactions on Fuzzy Systems*, 14(6):781–792.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Nava, J., Kosheleva, O., and Kreinovich, V. (2012). Why bernstein polynomials are better: fuzzy-inspired justification. In *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, pages 1–6. IEEE.

Nepali, R. K. and Wang, Y. (2013). Sonet: A social network model for privacy monitoring and ranking. In *2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops*, pages 162–166. IEEE.

Ni, Q., Bertino, E., Lobo, J., Brodie, C., Karat, C.-M., Karat, J., and Trombeta, A. (2010). Privacy-aware role-based access control. *ACM Transactions on Information and System Security (TISSEC)*, 13(3):24.

Nikolaenko, V., Ioannidis, S., Weinsberg, U., Joye, M., Taft, N., and Boneh, D. (2013). Privacy-preserving matrix factorization. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 801–812. ACM.

Obar, J. A. and Wildman, S. S. (2015). Social media definition and the governance challenge-an introduction to the special issue. *Telecommunications Policy*, 39(9):745–750.

Osatuyi, B. (2013). Information sharing on social media sites. *Computers in Human Behavior*, 29(6):2622–2631.

Palen, L. and Dourish, P. (2003). Unpacking privacy for a networked world. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 129–136. ACM.

Pensa, R. G. and Di Blasi, G. (2017). A privacy self-assessment framework for online social networks. *Expert Systems with Applications*, 86:18–31.

Petkos, G., Papadopoulos, S., and Kompatsiaris, Y. (2015). Pscore: A framework for enhancing privacy awareness in online social networks. In *Availability, Reli-*

*ability and Security (ARES), 2015 10th International Conference on*, pages 592–600. IEEE.

Preibusch, S. and Beresford, A. R. (2009). Privacy-preserving friendship relations for mobile social networking. In *W3C Workshop on the Future of Social Networking*, pages 1–5. Citeseer.

Raynes-Goldie, K. S. (2012). *Privacy in the age of Facebook: Discourse, architecture, consequences, Thesis*. Curtin University.

Renner, C. (2010). *Privacy in Online Social Networks*. Thesis.

Resnick, P. and Varian, H. R. (1997). Recommender systems. *Commun. ACM*, 40(3):56–58.

Ricci, F., Rokach, L., and Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer.

Rich, E. (1979). User modeling via stereotypes. *Cognitive Science*, 3(4):329–354.

Richthammer, C., Netter, M., Riesner, M., Sänger, J., and Pernul, G. (2014). Taxonomy of social network data types. *EURASIP Journal on Information Security*, 2014(1):11.

Rogers, A., Romanov, A., Rumshisky, A., Volkova, S., Gronas, M., and Gribov, A. (2018). Rusentiment: An enriched sentiment analysis dataset for social media in russian. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 755–763.

Rosenblum, D. (2007). What anyone can know: The privacy risks of social networking sites. *IEEE Security and Privacy*, 5(3):40–49.

Samarati, P. (2001). Protecting respondents identities in microdata release. *IEEE transactions on Knowledge and Data Engineering*, 13(6):1010–1027.

Samarati, P. and Sweeney, L. (1998). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Report, Technical report, SRI International.

Schilling, R. L., Song, R., and Vondracek, Z. (2012). *Bernstein functions: theory and applications, Book*, volume 37, ISBN:978-3-11-025229-3. Walter de Gruyter.

Schütze, H., Manning, C. D., and Raghavan, P. (2008). *Introduction to information retrieval*, volume 39. Cambridge University Press.

Settles, B. (2004). Biomedical named entity recognition using conditional random fields and rich feature sets. In *Proceedings of the international joint workshop on natural language processing in biomedicine and its applications*, pages 104–107. Association for Computational Linguistics.

Shang, S., Hui, Y., Hui, P., Cuff, P., and Kulkarni, S. (2014). Beyond personalization and anonymity: Towards a group-based recommender system. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, pages 266–273. ACM.

Shokri, R., Pedarsani, P., Theodorakopoulos, G., and Hubaux, J.-P. (2009). Preserving privacy in collaborative filtering through distributed aggregation of offline profiles. In *Proceedings of the third ACM conference on Recommender systems*, pages 157–164. ACM.

Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3):477–564.

Sramka, M. (2015). *Evaluating Privacy Risks in Social Networks from the User's Perspective, Book*, pages 251–267. Springer.

Srivastava, A. and Geethakumari, G. (2013). Measuring privacy leaks in on-

line social networks. In *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on*, pages 2095–2100. IEEE.

Stan, J., Muhlenbach, F., and Largeron, C. (2014). Recommender systems using social network analysis: Challenges and future trends. In *Encyclopedia of Social Network Analysis and Mining*, pages 1522–1532. Springer.

Stone, E. F., Gueutal, H. G., Gardner, D. G., and McClure, S. (1983). A field experiment comparing information-privacy values, beliefs, and attitudes across several types of organizations. *Journal of applied psychology*, 68(3):459.

Sun, X., Li, M., and Wang, H. (2011). A family of enhanced (l, $\alpha$)-diversity models for privacy preserving data publishing. *Future Generation Computer Systems*, 27(3):348–356.

Sun, X., Wang, H., Li, J., and Truta, T. M. (2008). Enhanced p-sensitive k-anonymity models for privacy preserving data publishing. *Trans. Data Privacy*, 1(2):53–66.

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570.

Taheri, S., Hartung, S., and Hogrefe, D. (2010). Achieving receiver location privacy in mobile ad hoc networks. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pages 800–807. IEEE.

Talukder, N., Ouzzani, M., Elmagarmid, A. K., Elmeleegy, H., and Yakout, M. (2010). Privometer: Privacy protection in social networks. In *Data Engineering Workshops (ICDEW), 2010 IEEE 26th International Conference on*, pages 266–269. IEEE.

Tambe, P. and Vora, D. (2016). Data sanitization for privacy preservation on social

network. In *Automatic Control and Dynamic Optimization Techniques (ICACDOT), International Conference on*, pages 972–976. IEEE.

Terrovitis, M., Mamoulis, N., Liagouris, J., and Skiadopoulos, S. (2012). Privacy preservation by disassociation. *Proceedings of the VLDB Endowment*, 5(10):944–955.

Thelwall, M. (2008). Social networks, gender, and friending: An analysis of myspace member profiles. *Journal of the Association for Information Science and Technology*, 59(8):1321–1330.

Torres-García, A. A., Reyes-García, C. A., Villaseñor-Pineda, L., and García-Aguilar, G. (2016). Implementing a fuzzy inference system in a multi-objective eeg channel selection model for imagined speech classification. *Expert Systems with Applications*, 59:1–12.

Toutanova, K., Klein, D., Manning, C. D., and Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 173–180. Association for Computational Linguistics.

Trewin, S. (2000). Knowledge-based recommender systems. *Encyclopedia of library and information science*, 69(Supplement 32):180.

Tuunainen, V. K., Pitkänen, O., and Hovi, M. (2009). Users' awareness of privacy on online social networking sites-case facebook. In *Bled 22nd e conference*, volume 42, pages 42–59.

Veiga, M. H. and Eickhoff, C. (2016). Privacy leakage through innocent content sharing in online social networks. *arXiv preprint arXiv:1607.02714*.

Walden, I. (2002). Anonymising personal data. *International JL & Info. Tech.*, 10(2):224–255.

Walters, C. (2009). Facebook's new terms of service:" we can do anything we want with your content. forever.". Technical report, The Consumerist.

Wang, D., Zeng, X.-J., and Keane, J. A. (2013). A simplified structure evolving method for mamdani fuzzy system identification and its application to high-dimensional problems. *Information Sciences*, 220:110–123.

Wang, K., Fung, B. C. M., and Philip, S. Y. (2007). Handicapping attacker's confidence: an alternative to k-anonymization. *Knowledge and Information Systems*, 11(3):345–368.

Wang, K., Yu, P. S., and Chakraborty, S. (2004). Bottom-up generalization: A data mining solution to privacy protection. In *Data Mining, 2004. ICDM'04. Fourth IEEE International Conference on*, pages 249–256. IEEE.

Wang, W., Chen, L., and Zhang, Q. (2015). Outsourcing high-dimensional healthcare data to cloud with personalized privacy preservation. *Computer Networks*, 88:136–148.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics.

Wittes, B. and Kohse, E. (2017). *The privacy paradox II: Measuring the privacy benefits of privacy threats, Book*. Center for Technology Innovation at Brookings, Washington DC.

Wondracek, G., Holz, T., Kirda, E., and Kruegel, C. (2010). A practical attack

to de-anonymize social network users. In *Security and Privacy (SP), 2010 IEEE Symposium on*, volume 1, pages 223–238.

Wong, R. C.-W., Li, J., Fu, A. W.-C., and Wang, K. (2006). ($\alpha$, k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 754–759. ACM.

Xiao, X. and Tao, Y. (2006a). Anatomy: Simple and effective privacy preservation. In *Proceedings of the 32nd International Conference on Very Large Data Bases*, pages 139–150. VLDB Endowment, ACM.

Xiao, X. and Tao, Y. (2006b). Personalized privacy preservation. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 229–240. ACM.

Xiao, X. and Tao, Y. (2007). M-invariance: towards privacy preserving re-publication of dynamic datasets. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 689–700. ACM.

Xu, H., Gupta, S., Rosson, M. B., and Carroll, J. M. (2012). Measuring mobile users' concerns for information privacy. In *Information Systems Security and Privacy*, pages 1–16. ICIS proceedings.

Xu, J., Wang, W., Pei, J., Wang, X., Shi, B., and Fu, A. W.-C. (2006). Utility-based anonymization using local recoding. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 785–790. ACM.

Xu, K., Guo, Y., Guo, L., Fang, Y., and Li, X. (2017). My privacy my decision: Control of photo sharing on online social networks. *IEEE Transactions on Dependable and Secure Computing*, 14(2):199–210.

Zeadally, S. and Badra, M. (2015). *Privacy in a Digital, Networked World, Book*. Springer, ISBN.978-3-319-08469-5, pp.1–417.

Zelenko, D., Aone, C., and Richardella, A. (2003). Kernel methods for relation extraction. *Journal of machine learning research*, 3(Feb):1083–1106.

Zhang, L., Li, X.-Y., Lei, J., Sun, J., and Liu, Y. (2015a). Mechanism design for finding experts using locally constructed social referral web. *Parallel and Distributed Systems, IEEE Transactions on*, 26(8):2316–2326.

Zhang, L., Li, X.-Y., Liu, K., Jung, T., and Liu, Y. (2015b). Message in a sealed bottle: Privacy preserving friending in mobile social networks. *IEEE Transactions on Mobile Computing*, 14(9):1888–1902.

Zhang, L., Li, X.-Y., and Liu, Y. (2013). Message in a sealed bottle: Privacy preserving friending in social networks. In *Distributed Computing Systems (ICDCS), 2013 IEEE 33rd International Conference on*, pages 327–336. IEEE.

Zheleva, E. and Getoor, L. (2009). To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *Proceedings of the 18th international conference on World wide web*, pages 531–540. ACM.

Zheleva, E., Terzi, E., and Getoor, L. (2012). Privacy in social networks. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 3(1):1–85.

# Appendix A
# Data Gathering Guide & Survey

The accompanying section shows the selection and recruitment methods of participants in this study. This section also provides the survey questions which were provided for the participants to answer. This information has been used in the evaluation section of Chapter 4 of this thesis.

## A.1 Participants Selection and Recruitment

As mentioned, the aim of this thesis is to investigate methods and techniques to measure and preserve the users' privacy when sharing their information in online social networks. To achieve this, there is a need to recruit potential participants who meet the study selection criteria and requirements. For contacting and recruiting participants of this part of the study, an email was sent to each participant including an information sheet and an online survey that detailed what the study was. "Purposive" or "convenience" sampling for data gathering was used. Convenience sampling will enable a comparable distribution of individuals in both groups of social networks. By applying this method, members of the community who best fit the study's requirements will be selected. In addition, no exclusive collection method that may exclude any gender, ethnic minorities and so on will be used.

## A.2 Information Sheet & Consent Form

The information sheet of this study is as follow:

Dear participant: you are invited to participate in a research study that aims to explore the methods of measuring users' privacy and anonymising sensitive information in online social networks. This study will be undertaken by Erfan Aghasian under the supervision of Dr Saurabh Garg and Dr James Montgomery at the University of Tasmania, in partial fulfilment of a PhD degree.

As online social sites facilitate data sharing and communication between individuals, they bring a number of risks, such as potential security breaches. Meanwhile, these sites ask users to share attributes such as age, name, interests, job details and so on. In this regard, evaluating users' privacy and anonymising their information are among the most significant challenges that should be considered. The evaluation of the sensitivity level of information (how much an attribute is important for a user in social media and how concerned you would be if that attribute were known publicly) can increase the users' awareness about their privacy. The anonymisation method can help to safeguard the users' information and enable users to understand which types of information have higher impact on their privacy.

You have been invited to participate in this study because you are between 25 and 50 years of age and participate in at least two different (Facebook and Twitter) social networks. The study involves collecting the level of sensitivity of users' attributes as well as the attributes value of users (data which is publicly available) which are shared on online social networks for providing privacy preservation methods for users. Involvement in this research carries negligible risk as the actual values of your publicly shared data are not shared with any other party and will be stored on a secure machine located in the School of Technology, Environments and Design at the University of Tasmania, Australia. It is unlikely that the data could be obtained by an external party. So, participants will not be at risk of

phishing, identity theft or other security attacks.

The study will provide you with a better understanding and insight about your privacy level. Hence, you can consider mitigation actions if you feel that your privacy level is not at an acceptable level. Moreover, the anonymisation technique will help you know what attributes can or cannot bring a privacy risk for you and what types of information you can safely share for the purpose of friending in online social networks.

If you wish to discuss any aspect of this study, please feel free to contact the researcher, Erfan Aghasian via email (erfan.aghasian@utas.edu.au) or phone (+61 3 6226 7897). You can also contact the student supervisor, Dr. Saurabh Garg via email (Saurabh.garg@utas.edu.au) or phone (+61 36226 6210).

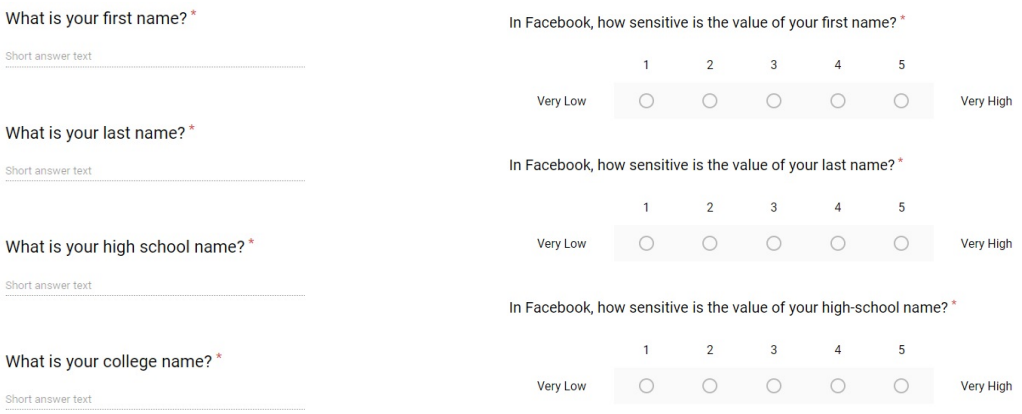
By submitting this form you agree that:

1. The nature and possible effects of the study have been explained clearly.
2. You understand that the study involves using your information from multiple social networks sites in an anonymised manner.
3. You understand that all research data will be securely stored on the University of Tasmania premises for five years from the publication of the study results and will then be destroyed.
4. Any questions that you have asked, should be answered to your satisfaction
5. You understand that the researcher(s) will maintain confidentiality and that any information supplied to the researcher(s) will be used only for the purposes of the research.
6. You understand that the results of the study will be published such that you cannot be identified as a participant.
7. You understand that your participation is voluntary and that I may withdraw at any time without any effect.

This study has been approved by the Tasmania Social Sciences Human Research Ethics Committee. If you have concerns or complaints about the conduct of this study, please contact the Executive Officer of the HREC (Tasmania) Network on (+61 3 6226 6254) or email (human.ethics@utas.edu.au). The Executive Officer is the person nominated to receive complaints from research participants.

## A.3 Survey Questions

In order to gather the data of participants, users have been asked what they have stored publicly in their Facebook and Twitter profiles. For the structured data, 16 attributes for Facebook and 7 attributes for Twitter were considered. (These are the attributes that has been used in previous research studies (Srivastava and Geethakumari, 2013; Gross and Acquisti, 2005) of online social network topics). Facebook attributes include first name, last name, high school, faculty, university, hometown, state, gender, date of birth, zip-code, current job, relationship status, qualifications, interests, religious views and political views. Twitter attributes include first name, last name, profile description, location, website address and date of birth. For unstructured data, participants were asked to provide the last ten publicly visible posts or comments from Facebook and the last ten tweets from Twitter.Then, users were asked to enter the publicly-visible values they share in their Facebook and Twitter profiles in the designated section. (*For example, what is your first name?*). If they had not recorded an attribute in Facebook or Twitter, or it is not shared publicly, then they were asked to enter the text 'private'. After asking the level of visibility for the information, the sensitivity of the attributes was questioned. Five different levels of sensitivity have been assigned, including 1 as very low sensitivity, 2 as low sensitivity, 3 as medium sensitivity, 4 as high sensitivity and 5 as very high sensitivity. Then, participants were asked *how sensitive is the value of each attribute is?*




(a) Attribute Visibility (b) Attribute Sensitivity

Figure A.1: Example of Survey's Questions

Finally, participants were asked to provide the ten most recent public posts from Facebook and the ten most recent tweets from Twitter to check the credibility of the proposed model. Figures A.1a and A.1b show some example questions that have been asked of the users.